Structural Insight into the Catalytic Mechanism of the Unique Glutathione-Dependent Peroxidase Enzymes of *Trypanosoma Cruzi*

Shreenal Patel

Institute of Structural & Molecular Biology

University College London

A thesis submitted for the degree of

Doctor of Philosophy

September 2008

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Statement

All work presented in this thesis was carried out by myself. Enzyme kinetics was carried out together with Dr. Syeed Hussain and structural refinement in CNS was carried out by Dr. Snezana Djordjevic. All NMR spectra presented was collected and processed by Dr. Richard Harris.

I, Shreenal Patel, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Kinetoplastid parasites cause a variety of diseases including African sleeping sickness (*Trypanosoma brucei*) and Chagas disease (*Trypanosoma cruzi*). The invading parasite is exposed to reactive oxygen species (ROS) generated from the host immune response. In trypanosomes, the glutathione/glutathione reductase antioxidant system (present in the host), is replaced by the analogous trypanosome-specific trypanothione/trypanothione reductase antioxidant system. The differences in the unique oxidative defence system of the parasite from its host, provides potential drug targets. The parasites, like their mammalian hosts, possess glutathione-dependent peroxidase enzymes (GPX), which have a conserved active site cysteine (analogous to selenocysteine in the mammalian hosts). Research now considers these enzymes to fall into a mechanistic family similar to peroxiredoxin enzymes. Peroxiredoxins are a large family of proteins, whereby members fall into the subgroups according to the number of active site cysteine residues present and their distribution (1-Cys or 2Cys typical/atypical). It was recently found that *T. brucei* glutathione peroxidase enzyme (TbPxIII) follows a mechanism analogous to atypical 2-Cys peroxiredoxin mechanism, supporting a similar mechanism in the homologue *T. cruzi* glutathione peroxidase I (TcGPXI).

The aim of my thesis work was to gain structural information on TcGPXI, to provide insight into the mechanism of this enzyme. Constructs of TcGPXI, TcGPXII, TbPxI and TbPxII were cloned, expressed and purified to produce sufficient protein for structural studies. TcGPXI was investigated by NMR spectroscopy, which facilitated 96% backbone sequence assignments and secondary structure prediction for oxidised wild type TcGPXI. An intra-molecular disulphide bond was observed between the conserved catalytic cysteine and a second cysteine within the protein sequence. To investigate this, cysteine mutants were created to break the disulphide bond and investigated with NMR techniques. Additionally, sequence-specific backbone assignments were obtained for reduced wild type TcGPXI. Established kinetic assays were used to assess catalytic activity. The mutant enzymes exhibited no activity, suggesting that both cysteines involved in the intramolecular disulphide bond, are required for catalysis. Protein from constructs of TcGPXII, TbPxI, TbPxII and oxidised wild type TcGPXI (including mutant forms) were subjected to crystallisation screens. Diffraction data for oxidised wild type TcGPXI was

collected, at a resolution of 2.3 Å. Molecular replacement was used to solve the crystal structure of TcGPXI.

The work presented in this thesis will show that TcGPXI follows a new mechanistic family of enzymes, similar to that of the atypical 2-Cys peroxiredoxin family. The crystal structure reveals the possibility of a disulphide bond in the oxidized state, associated with the unravelling of a helix otherwise present in the thioredoxin fold.

Acknowledgments

I would firstly like to thank my supervisor, Dr. Snezana Djordjevic for the opportunity to carry out the work presented in this thesis and for all her encouragement and support through this work. I would like to thank her for her guidance in the unravelling of a journey from which I have benefited immensely not only through science but within myself.

Thanks to Professor Paul Driscoll for his support, those in my group and other laboratories who have provided encouragement and help along the way. I would like to especially express my gratitude to Dr. Syeed Hussain who has played a prominent role through the course of my studies with endless encouragement, support, and enthusiasm when morale has been low! I am a firm believer in "the story" now! Thanks to Dr. Richard Harris who has given me so much of his time and been invaluable throughout my work in providing insightful discussions. I thank both Richard and Syeed for their pearls of wisdom and patience with me! Thanks to Professor Stephen Perkins for his support and use of computer facilities, and Dr. Gary Parkinson for all his help with crystallography work.

I could not have got through the last years without the constant support and care of my closest friends and family. In particular, I would like to thank Dr. Alex Bonner and Dr. Becky Saunders for their endless support – they have been a backbone to me. I feel honoured to have the friendships and love that I have received from them both. I am the last of us three, but look forward to sharing many more milestones together. A special thank you to Alex – I am so grateful to you for your constant support, care and faith in me. Thanks to Pati and Hannah for all their encouragement and laughter filled evenings.

Last but by no means least I would like to thank my parents who came to this country with nothing and have struggled so hard for so many years to provide a better future for me and others. Words cannot express how grateful I am to you both for giving me so much opportunity in all aspects of life. If I can be even a fraction of the people that you both are then I would be more than happy. You are both heroes. Thanks to all my family for support and especially to Ronnie and my new family for so much encouragement and care.

I dedicate this work to my parents and in loving memory of my grandfather who is always greatly missed, C. B. Patel.

Contents **Page**

Contents - Figures **Page**

Chapter One

Chapter Two

Chapter Three

Contents - Figures **Page**

Chapter Four

Contents - Figures **Page**

Contents - Figures **Page**

Chapter Five

Contents - Tables **Page**

Chapter Two

Chapter Three

Chapter Four

Chapter Five

Appendix

Abbreviations

Å	angstroms ($\times 10^{-13}$ meter)
AIDS	acquired immunodeficiency disorder
APX	ascorbate-dependent haemoperoxidase
Asc	ascorbate
BCSB	Bloomsbury Centre for Structural Biology
bp	base pair
BSA	bovine serum albumin
C. fasciculata	*Crithidia fasciculata*
CfTR	*Crithidia fasciculata* trypanothione reductase
CIP	calf intestinal phosphatase
CSI	chemical shift index
CPX	cytosolic peroxiredoxin
C_P	peroxidatic cysteine/selenocysteine
C_R	resolving cysteine
1D	one dimensional
2D	two dimensional
3D	three dimensional
DHA	dehydroascorbate
DNA	deoxyribonucleic acid
dNTP	deoxynucleotide triphosphate
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dTTP	deoxythymidine triphosphate
DTT	dithiothreitol
EB	elution buffer
E. coli	*Escherichia coli*
EDTA	ethylenediaminetetraacetic acid
EGF	epidermal growth factors
FAD	flavin adenine dinucleotide
FID	free induction decay

FPLC	fast protein liquid chromatography
GPX	glutathione-dependent peroxidases
GSH	glutathione
GSSG	oxidised glutathione
GR	glutathione reductase
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	heteronuclear single quantum coherence
IPTG	isopropyl-β-D-thiogalactopyranoside
K	Kelvin
° C	degrees Celsius
kb	kilobases
kDa	kiloDaltons
LB	Luria-Bertani
LDL	low-density lipoproteins
LSHTM	London School of Hygiene and Tropical Medicine
L. major	*Leishmania major*
L	litre
M	molar
ml	millilitre
mg	milligram
µg	microgram
µl	microlitre
MHz	megahertz
MPX	mitochondrial peroxiredoxin
ms	milliseconds
NADPH	nicotinamide adenine dinucleotide phosphate (reduced form)
NADP	nicotinamide adenine dinucleotide phosphate
Ni-NTA	nickel ion (Ni^{2+}) nitriloacetic acid histidine binding resin
NEB	New England Biolabs
nm	nanometer ($\times 10^{-9}$ meter)
NMR	nuclear magnetic resonance
OD_{600}	optical density at 600 nanometers
PCR	polymerase chain reaction

PDB	protein data bank
PDGF	platelet-derived growth factor
PEG	polyethylene glycol
PHGPX	phospholipid GPX
ppm	parts per million
Prx	peroxiredoxin
PtGPX5	*Populus trichocarpa x deltoides* GPX5
R_{factor}	residual factor
R_{free}	free residual factor
rcf	relative centrifugal force
rpm	revolutions per minute
RMSD	root mean square deviation
ROS	reactive oxygen species
RNA	ribonucleic acid
SCOP	structural classification of proteins
SDS-PAGE	sodium dodecyl sulphate-polyacrylamide gel electrophoresis
SEC	size exclusion chromatography
TAE	tris-acetate- ethylenediaminetetraacetic acid
T. brucei	*Trypanosoma brucei*
Tb	*Trypanosoma brucei*
TbPxI	*Trypanosoma brucei* GPX-type tryparedoxin peroxidase I
TbPxII	*Trypanosoma brucei* GPX-type tryparedoxin peroxidase II
TbPxIII	*Trypanosoma brucei* GPX-type tryparedoxin peroxidase III
T. cruzi	*Trypanosoma cruzi*
Tc	*Trypanosoma cruzi*
TEMED	tetramethylethylenediamine
TEV	*Tobacco Etch virus*
Tris	tris(hydroxymethyl)aminomethane
TrxR	thioredoxin reductase
Trx	thioredoxin
TR	trypanothione reductase
$T[SH]_2$	dihydrotrypanothione
TS_2	trypanothione
TXN/TPN	tryparedoxin

UCL	University College London
UMSBP	universal minicircle sequence-binding protein
UV	ultra-violet
V	volts
VSG	variant surface glycoproteins

Amino Acid Abbreviations

Amino Acid	Three Letter Code	One Letter Code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartic Acid	Asp	D
Cysteine	Cys	C
Glutamine	Gln	Q
Glutamic Acid	Glu	E
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Selenocysteine	SeCys	U
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Chapter One

INTRODUCTION

1.1 Kinetoplastids

The Kinetoplastids are a group of flagellated protozoan parasites and are pathogens of a range of invertebrates, vertebrates and some plants (Simpson et al., 2002). The kinetoplastids are characterised by the presence of the kinetoplast organelle, containing their single large mitochondrial DNA, which is attached to the basal bodies of the flagellum and undergoes repression and activation so that the trypanosome parasite is able to switch its pattern of respiration to match the host energy source (Ploubidou et al., 1999).

Kinetoplastid protozoan parasites cause distinct human diseases, the most common of which are human African trypanosomiasis (sleeping sickness) caused by various subspecies of *Trypanosoma brucei* (*T. brucei*; Tb), Chagas disease caused by *Trypanosoma cruzi* (*T. cruzi*; Tc) and numerous forms of Leishmaniasis caused by different forms of the species *Leishmania* (Stuart et al., 2008). There are a large number of similarities between the various parasitic members of the genus *e.g.* all kinetoplastids are motile with a single flagellum that occurs from a pocket in the cell membrane (where endocytosis also occurs); their peroxisomes are modified to perform glycolysis and thus termed glycosomes; they typically grow in an asexual manner and divide by binary fission during which their nucleus does not undergo membrane dissolution or chromosome condensation.

The genomes of *T. cruzi*, *T. brucei* and *Leishmania major* (*L. major*) have been sequenced (Berriman et al., 2005; El-Sayed et al., 2005a; Ivens et al., 2005), where the haploid genomes consist of 25-55 megabases distributed over 11-36 mostly diploid chromosomes that encode 8100 (*T. brucei*), 8300 (*L. major*) and 12,000 (*T. cruzi*) protein-coding genes. The genomes exhibit a high level of syntenic conservation shown by a conserved core proteome of approximately 6200 genes organised in large syntenic polycistronic gene clusters, from which it is estimated that these organisms diverged 200-500 million years ago and have evolved to differ in their vectors, mechanism of immune evasion, target tissue and pathogenesis (Kissinger, 2006; El-Sayed et al., 2005b).

1.2 Chagas disease

The trypanosomal parasite *T. cruzi*, causative agent of Chagas disease, is considered to be the most significant disease-causing parasite in Latin America, where up to 20 million people carry the infection of which one third will die, making it a major health problem in areas of economic burden (Wilkinson et al., 2002b). The complex epidemiology of the disease is not fully understood however *T. cruzi* has been shown to exhibit significant genetic diversity and hybridisation in various sub-groups of the species. Until recently it was considered that genetic exchange was absent in *T. cruzi*. However population genetic analyses now suggest that genetic exchange occurred during the early evolution of these parasites, accounting for the high degree of genetic diversity in the species, although the phenomenon is considered to be possibly absent in present *T. cruzi* populations. It has subsequently been suggested that distinct *T. cruzi* genotypes cause various degrees of severity in forms of Chagas disease (Teixeira et al., 2006; Kelly, 2003).

The lifecycle of the *T. cruzi* parasite is diagrammatically represented in Figure 1.1. The main source of transmission of the parasite is by the triatomine insect, which transmits infective metacyclic trypomastigote forms of the parasite to the human host via a blood meal or via its faeces. These parasitic forms are then able to pass into the body of the host through the scratching of the insect bite or through permissive mucosa or conjuctival membranes. Inside the mammalian host the *T. cruzi* parasite is predominantly intracellular and does not multiply in the bloodstream – instead they invade non-phagocytic or phagocytic cells and divide by binary fission as small ovoid amastigotes, which do not have a flagellum. Pseudocysts may form in various tissues that rupture after five days to release new mobile trypomastigotes, some of which can re-invade cells whilst others circulate in the bloodstream to infect new sites. The vector stage of the parasitic lifecycle commences when the triatomine bug ingests a bloodmeal from the infected mammalian host, where trypomastigotes transform into epimastigotes in the insect midgut, proceeding into stages of multiplication and differentiation into metacyclic trypomastigote forms. In addition to the transmission by the insect vector, the parasite can also be transmitted through blood transfusions, organ transplantation, transplacentally and from mother to child via breast milk.

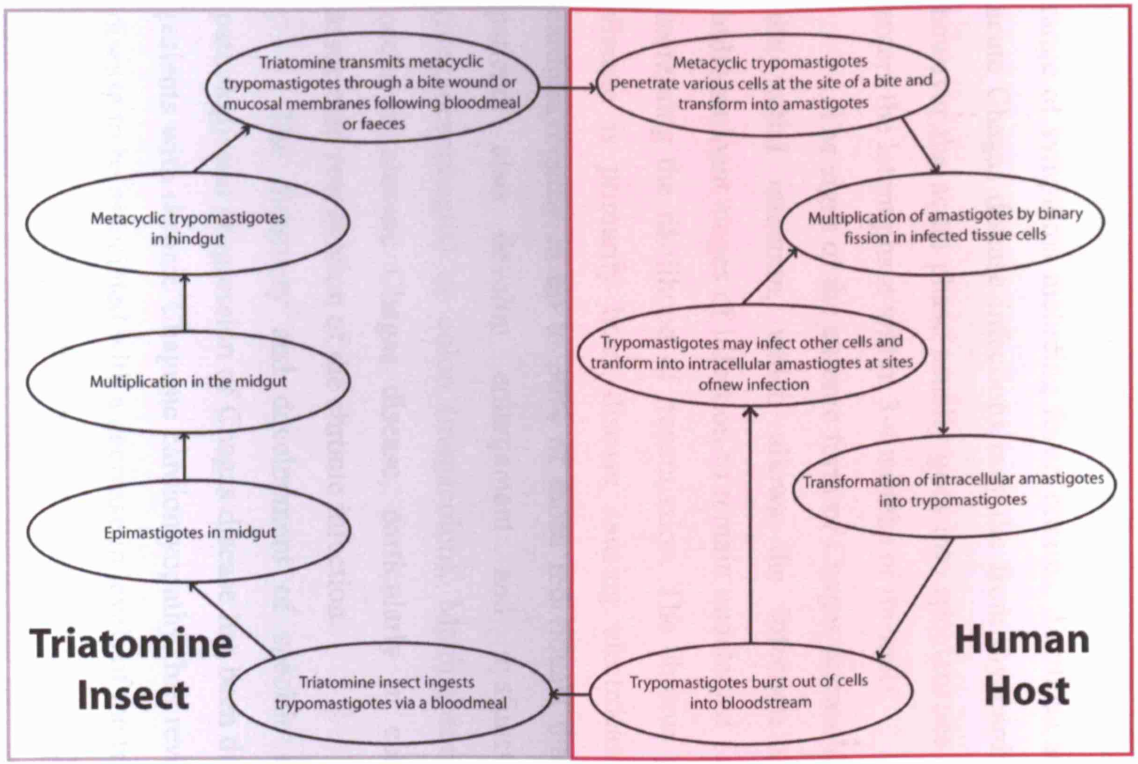


Figure 1.1: Lifecycle of *Trypanosoma cruzi*

A diagrammatic representation of the alternating lifecycle stages of the *T. cruzi* parasite in the human host and the triatomine insect vector.

(Adapted from Division of Parasitic Diseases Centre for Disease Control and Prevention: http://www.cdc.gov/ncidod/dpd/index.htm)

Although the pathogenesis of Chagas disease is only partially understood the clinical course of Chagas disease has three stages: the acute phase (3-4 months); indeterminant phase (10-30 years latency); and the chronic phase (symptoms include cardiomyopathy and congestive heart failure).

In the acute phase of Chagas disease, 95% of cases are asymptomatic and thus not perceived by patients. The symptomatic cases are characterised by a wide range of symptoms including fever, cramps, diarrhoea and comatosis. Patients with acute Chagas disease infections may die from myocarditis or meningoencephalitis, however the acute phase usually goes into spontaneous remission and the infection enters the latent phase within 3-4 months of its onset.

The signs of the chronic form of Chagas disease usually arise only years after the initial infection, which allows the infection of those individuals with indeterminant stages of infection to remain undetected for long periods of time, thus increasing the likelihood of transmission. The chronic phase symptoms of Chagas disease is primarily heart disease causing electrocardiogram abnormalities and cardiomyopathy in up to 30% of those individuals that are affected, where some patients also develop enlargement and dysfunction of the oesophagus (megaoesophagus) or colon (megacolon). Meningoencephalitis is also shown to occur in chronic Chagas disease, particularly in congenital cases and AIDS-associated reactivation of the chronic infection.

The discovery and development of specific markers for following the pathology and progression of Chagas disease has been difficult. Extensive studies on patients with chronic Chagasic cardiomyopathy, has revealed the progression of the disease to be associated with a decrease in levels of the trace element selenium which could be utilised as a biological marker (Rivera et al., 2002). Another factor that has been studied in an attempt to follow the disease pathology and its diagnosis is the decrease in the activity of mitochondrial respiratory complexes in Chagasic patients using the oxidative damage marker malonylaldehyde. Additionally, these patients exhibited a decrease in glutathione peroxidase enzymes, superoxide dismutase and glutathione. Patients were shown to exhibit sustained oxidative damage, which would increase the amount of free radicals, and in turn lead to further increased oxidative damage (Wen et al., 2006).

There are no current vaccinations against Chagas disease, thus the use of drug treatments is paramount for the treatment and management of the disease. For more

than forty years common drug therapies have utilised the nitroheterocyclic drugs nifurtimox and benznidazole, which function by enzyme-mediated activation within the parasite by decreasing thiol levels within the parasite. This increases toxicity in the parasite due to an elevation in levels of reactive oxygen species which the parasitic antioxidant defence system is inefficient in dealing with (Wilkinson et al., 2008). The action of these drugs has been found to have limitations *e.g.* drug therapies are highly dependent on the phase of the disease, the susceptibility of the *T. cruzi* strain itself (some strains may confer resistance) and the physiological conditions of the host. It has also been reported that these drugs show poor efficiency in preventing the progress of Chagas disease in the chronic phase. Such inefficiencies, toxicity of the chemicals and the harmful side-effects highlight the requirement for new targets in combating Chagas disease (Andrade et al., 2008; Murta et al., 1998).

1.3 Sleeping sickness

Human African trypanosomiasis (sleeping sickness) occurs in the poorest rural populations in some of the least developed countries in sub-Saharan Africa. The disease is caused by the trypanosomal parasite *T. brucei*, of which there are two forms of *T. brucei*: *T. brucei gambiense* (causes chronic condition with months or years of a passive phase) which is found in central and western Africa and *T. brucei rhodesiense* (acute form that is more virulent and develops faster) found in southern and eastern Africa. Other sub-species *T. brucei brucei* and *T. brucei congolense* are the causative agent of African trypanosomiosis in animals (Nagana cattle disease).

The completed genome sequence of *T. brucei* reveals the presence of 11 megabase-sized chromosomes with unspecified numbers of smaller and intermediate-sized chromosomes (30-700 kb) known to encode sequences similar to subtelomeric regions of the megabase-sized chromosomes (26-megabase genome). Whilst *T. cruzi* is considered to have a clonal population structure, *T. brucei* has recently been suggested to undergo frequent sexual recombination and genetic exchange (MacLeod et al., 2001; Katz L.A. and Bhattacharya D., 2006), however the populations infecting humans (*T. brucei gambiense*, *T. brucei rhodesiense)* are shown to exhibit clonal population structures. The importance of genetic diversity to

the pathogenesis of the parasite is shown in the evasion of the host immune system by the *T. brucei* parasite, through switching of a group of variant surface glycoproteins (VSGs) where the importance of recombination is shown through the movement of silent VSG into active expression sites. Contrastingly in *T. cruzi* the use of antigenic strain variation is not used, instead the parasite relies on the high degree of population heterogeneity for adapting to the host environment (Machado et al., 2006). In both of these parasites a complete understanding of the nature of genetic diversity and relationship to molecular pathogenesis is essential in the combat against Chagas disease and sleeping sickness.

The *T. brucei* parasite is transmitted by the bite of an infected tsetse fly, shown in Figure 1.2, which injects metacyclic trypomastigotes into the tissue of the skin upon which the parasite can then enter the lymphatic system and bloodstream. The bloodstream trypomastigotes are able to replicate by binary fission and can be carried to other body parts such as the spleen, liver, endocrine system and heart. The insect vector becomes infected with bloodstream trypomastigotes upon taking a bloodmeal from an infected human host, where the parasite then tranforms into procyclic trypomastigote forms and multiplies in the insect midgut. After leaving the mid-gut, the parasite transforms into the epimastigote form and continues multiplication in the salivary gland of the fly.

An individual can be infected with sleeping sickness for months or years before the emergence of symptoms. Initial symptoms include headaches, fever and anaemia, as well as swollen lymph nodes in the neck (Winterbottom's sign). In the second stage, as the parasite crosses the blood brain barrier, the symptoms progress to severe headaches, insomnia and various sleep disorders, mental deterioration, comatosis and death. Progression of the disease occurs faster in humans that are infected with the *T. brucei rhodesiense* than those infected with the *T. brucei gambiense* form (Burchmore et al., 2002).

There are no vaccinations against sleeping sickness and prospects of developing a successful vaccination are poor due to the ability of the parasite to alter their surface coat through antigenic variation. Current drug therapies present problems of toxicity and deleterious side-effects as well as drug resistance. The present treatment of sleeping sickness includes the drugs suramin (for early stages of the rhodienese form), pentamidine (for early stages of gambienese form), melarsoprol (for late stages of both rhodiense and gambienese forms) and

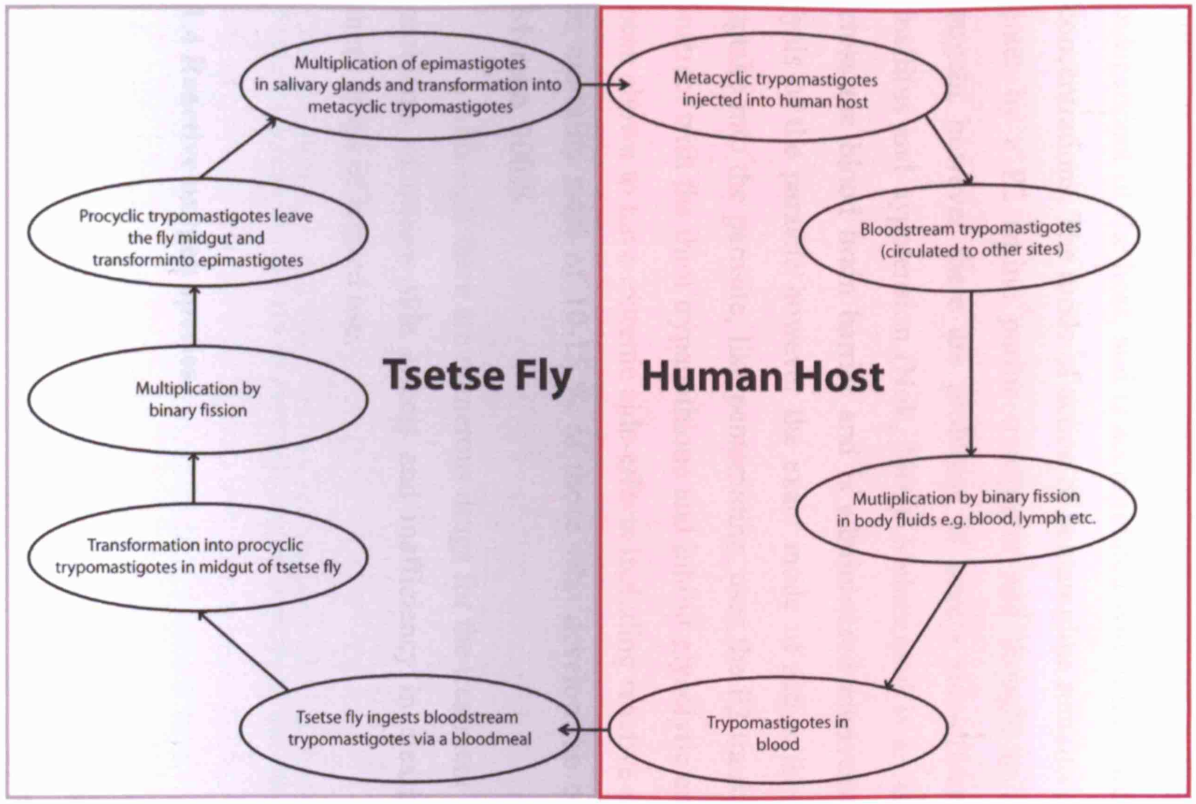


Figure 1.2: Lifecycle of *Trypanosoma brucei*

Diagrammatic representation of the alternating lifecycle stages of the *T. brucei* parasite in the human host and the tsetse fly insect vector. (Adapted from Division of Parasitic Diseases Centre for Disease Control and Prevention: http://www.cdc.gov/ncidod/dpd/index.htm)

eflornithine (for late stage of gambienese form only). The drug suramin is a napthalene derivative and thought to be accumulated by the parasite via endocytosis of the drug bound to low-density lipoproteins (LDL). The exact mode of action for suramin remains unclear although it has been suggested that the drug interferes with trypanosome metabolism by preventing the receptor-mediated uptake of LDL and is involved in the inhibition of enzymes present in glycosomes (Burchmore et al., 2002). Pentamidine is an aromatic diamidine, working directly against the parasite independent of the host, and is accumulated within the parasite reaching millimolar concentrations. The mode of action of pentamidine remains unclear but it is found to enter by a P2 amino purine transporter and thought to inhibit multiple cellular targets, however there are problems of severe side-effects that include Diabetes mellitus and hypotension (Nok, 2003). Melarsoprol is an arsenic derivative able to cross the blood brain barrier and is administered intravenously. The drug induces lysis of the parasite however the exact mode of action is yet to be established. Its uptake into the parasite, like pentamidine, uses the P2 transporters and is thought to interact with the thiol trypanothione and inhibit glycolytic enzymes. Melarsoprol has been shown to have extreme side-effects including reactive encephalopathy resulting in mortality rates of 10-15 % of those who develop the condition (Docampo and Moreno, 2003).

Although there are numerous drugs for the treatment of sleeping sickness, the problems of severe side effects and inefficiency in clearing parasitaemia, render these drugs of limited use.

1.4 Reactive oxygen species

The invading trypanosomal parasites and their human hosts are exposed to harmful reactive oxygen species (ROS). In aerobic organisms the reduction of molecular oxygen (O_2) under aerobic conditions usually occurs via a 4-electron transfer producing two molecules of water (H_2O). As a consequence of unavoidable bi-products of cellular respiration and the actions of some enzymes and redox molecules, molecular oxygen can be also reduced via 1 or 2- electron transfers generating reactive oxygen species (Figure 1.3). The products of this reduction

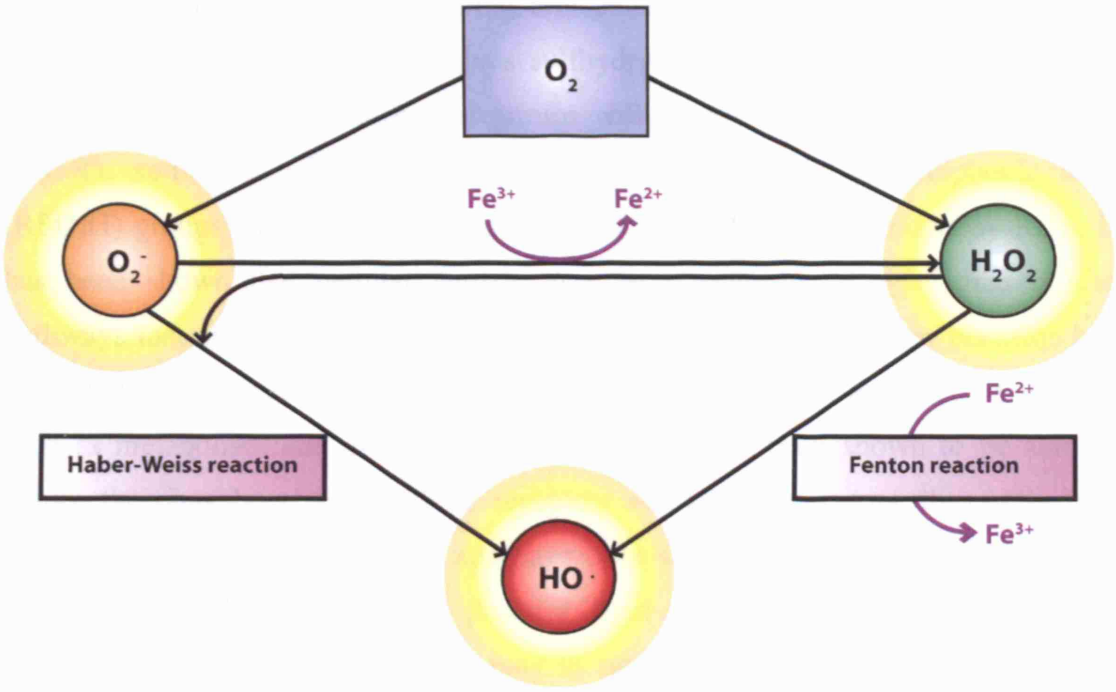


Figure 1.3: The formation of reactive oxygen species in the mammalian host

The generation of superoxide anions (O_2^-), hydrogen peroxide (H_2O_2) and hydroxyl radical ($HO^·$) reactive oxygen species (circled yellow) is shown from the incomplete reduction of molecular oxygen (O_2) and via the Fenton and Haber-Weiss reactions.

include superoxide anions (O_2^-) and hydrogen peroxide (H_2O_2). Hydrogen peroxide and superoxide anions are able to be reduced further by the Fenton reaction in the presence of Fe^{2+}, or by the Haber-Weiss reaction, to produce the most destructive and harmful ROS – the hydroxyl radical ($HO^{\cdot}$), which is able to cause DNA, protein and fatty acid damage. The damage of fatty acids generates lipid peroxides that can cause the destabilization of membranes.

ROS can also act as signalling molecules. The regulation of redox cellular signalling in most organisms involves sulfhydryl groups (R-SH) found on protein cysteine residues. These sulfhydryl groups will readily oxidise to give disulphide (RSSR), sulphenic acid (RSOH), sulphinic acid (RSO_2H) and sulphonic acid (RSO_3H) groups (Kamata and Hirata, 1999). When cells are stimulated by ROS, it has been shown that cellular signals may be transferred through some of the pathways that are triggered by growth factors. For example, hydrogen peroxide has been shown to trigger certain pathways that would normally be activated by insulin, such as metabolism and growth stimulation. ROS have been shown to be able to induce the phosphorylation of EGF receptors and PDGF receptors, thus activating downstream signalling pathways, and have been shown to be able to activate some serine/threonine kinases, small G proteins, lipid signalling and Ca^{2+} signalling. In addition to this, signalling systems may in turn regulate cellular redox states and increase ROS production. An example of the importance of ROS and redox signalling is shown in Alzheimer's and Parkinson's disease, where irregular redox cellular signalling contributes to the cell death (Kamata and Hirata, 1999).

In eukaryotic organisms the correct balance of ROS and prevention of their deficiency or accumulation causing oxidative stress is maintained through antioxidant mechanisms in various subcellular localisations involving enzymes such as superoxide dismutase (conversion of two superoxide anions into one molecule of hydrogen peroxide and one molecule of oxygen), catalase (conversion of two molecules of hydrogen peroxide into two molecules of water and one molecule of oxygen) and peroxidase enzymes.

1.5 Oxidative defence mechanisms

In various eukaryotic organisms the balance of reactive oxygen species at a non-toxic level is maintained by the action of antioxidant systems, of which the two major systems are glutathione/glutaredoxin pathway (includes glutathione, glutathione reductase and glutathione peroxidase enzymes) and the thioredoxin pathway (includes thioredoxin reductase and thioredoxin peroxidase enzymes).

Glutathione is a low-molecular weight thiol consisting of a tripeptide of glycine, cysteine and glutamate. The reduced form of glutathione is maintained at high levels in the cellular environment, which is essential for oxidative defence (Ottaviano et al., 2008). Glutathione plays an important role in the protection against cellular damage from excessive ROS, lipid peroxides and protein oxidation through various antioxidant pathways dependent on this thiol. In mammalian systems, the reduction of hydrogen peroxide and organic hydroperoxides is catalysed by the glutathione-dependent peroxidase enzymes, following the transfer of reducing equivalents from two molecules of glutathione. The resultant oxidised form of glutathione is then regenerated into the reduced form using the NADPH-dependent flavoenzyme glutathione reductase. Reduction of oxidised glutathione (GSSG) occurs by the transfer of electrons from NADPH to glutathione reductase (GR) via the FAD domain generating a dithiol (Cys58 and Cys63 in human GR). Subsequently, the binding of oxidised glutathione and formation of a mixed disulphide between Cys58 takes place, whereby one molecule of reduced glutathione is released. A second molecule of reduced glutathione is released by nucleophillic attack of Cys68 on the mixed disulphide intermediate which leads to the oxidised form of glutathione reductase (Pai and Schulz, 1983; Karplus and Schulz, 1989). The function of glutathione in antioxidant defence extends further to the transfer of reducing equivalents to two other possible proteins: the glutaredoxin proteins which can in turn reduce a variety of substrates including protein disulphides or glutathione-mixed disulphides (e.g. glutathionylated proteins), or to the glutathione S-transferases that catalyses the conjugation of glutathione with other molecules as an intermediate stage in detoxification (Fernandes and Holmgren, 2004).

A second molecule, thioredoxin, is a small (12 kDa) redox protein with an active site sequence of W-C-G-P-C, and participates in the reduction of oxidised proteins, hydroperoxides and redox regulation of apoptosis and cellular growth.

Mammalian thioredoxin proteins have a wide variety of functions including roles in antioxidant defence where reduced thioredoxin is able to transfer reducing equivalents to thioredoxin peroxidase enzymes (also commonly referred to as peroxiredoxin enzymes). Upon reduction of hydroperoxide species by thioredoxin peroxidases, oxidised thioredoxin is regenerated into its reduced form by the NADPH-flavoenzyme thioredoxin reductase (closely related to glutathione reductase).

Mammalian thioredoxin reductase isoenzymes are selenocysteine-containing homodimer oxidoreductases that contain one FAD molecule per subunit and are used to reduce oxidised thioredoxin. Mammalian thioredoxin reductase contains a conserved catalytic redox site at the sequence motif Cys-Val-Asn-Val-Gly-Cys (which is an identical sequence motif to that present in glutathione reductase) and additionally a selenocysteine (SeCys) containing C-terminus conserved sequence motif Gly-Cys-SeCys-Gly. Electrons flow from NADPH through FAD and reduce the thioredoxin reductase catalytic site generating a dithiol form (the C-terminus selenocysteine containing motif is oxidized at this point). The proceeding thiol-disulphide exchange where electrons are transferred from the catalytic site of one subunit to the C-terminal selenenylsulfide (Cys-SeCys) of the other subunit of the dimer, results in a reduction of the C-terminus site and the formation of a disulphide at the catalytic site. The reduced C-terminus site is then able to transfer reducing equivalents from thioredoxin reductase to thioredoxin and reduce oxidised thioredoxin (Nordberg and Arnθr, 2001; Mustacich and Powis, 2000). Despite the large number of widely researched similarities between the glutathione and thioredoxin systems, it appears that both are independently controlled and complementary to each other, thus providing an effective and powerful antioxidant defence system in many organisms including mammals (Watson et al., 2004).

The trypanosomal parasites *L. major*, *T. brucei* and *T. cruzi* have been found to possess thioredoxin proteins containing the conserved thioredoxin catalytic motif (W-C-G-P-C-K), but as yet no thioredoxin reductase has been found in these organisms or any other Kinetoplastid (Schmidt and Krauth-Siegel, 2003). Furthermore it has been shown that the *T. brucei* thioredoxin is able to be reduced by the trypanosome-specific thiol trypanothione and in turn reduce hydroperoxides via transfer of reducing equivalents to trypanosomal peroxiredoxin enzymes (Piattoni et al., 2006). Thioredoxins of *T. brucei* and *L. major* have been shown to form a distinct

phylogenetic branch in evolutionary analyses, where they are more closely related to mammalian thioredoxin than to yeast and plant thioredoxins.

Trypanosoma exhibit major differences from their mammalian hosts in their antioxidant defence systems that include: the presence of iron-containing superoxide dismutase (in mammals superoxide dismutase is a copper or zinc containing enzyme), the absence of the enzyme catalase, the absence of selenium-dependent glutathione peroxidase enzymes (in trypanosomal glutathione peroxidases selenocysteine is replaced by cysteine) and most importantly the use of a totally different thiol dependent antioxidant system. It is such differences in the distinct oxidative defence systems of the parasite and its host that provide potential drug targets.

1.6 Antioxidant systems in trypanosoma

The trypanosomal antioxidant defence system, as shown in Figure 1.4a, is dependent on their main low molecular weight thiol, trypanothione. Trypanothione has been found to be able to spontaneously reduce hydrogen peroxide and detoxify nitrosothiols but with lower activity than when coupled to the function of peroxidase enzymes (Castro and Tomas, 2008). In the oxidative defence pathway of *T. cruzi*, trypanothione is maintained in its reduced form (dihydrotrypanothione) by reducing equivalents from the NADPH-dependent flavoenzyme trypanothione reductase. Dihydrotrypanothione subsequently provides reducing equivalents to hydroperoxides via three possible pathways involving either: glutathione (GSH) and glutathione-dependent peroxidases (GPX); tryparedoxin (TXN, thioredoxin-like) and peroxiredoxins (Prx); or ascorbate (Asc) and ascorbate-dependent haemoperoxidase (APX), whose substrate specificities and localisations are shown in Figure 1.4b.

Trypanothione is formed from the conjugation of glutathione and spermidine (N^1, N^8 –bis (glutathionyl) spermidine) by the catalysis of the enzymes glutathionylspermidine synthetase and trypanothione synthetase (Figure 1.5). In the African trypanosome spermidine is formed from ornithine via putrescine and then used to conjugate glutathione, leading to the production of glutathionylspermidine and subsequently trypanothione. In the American trypanosome, trypanothione

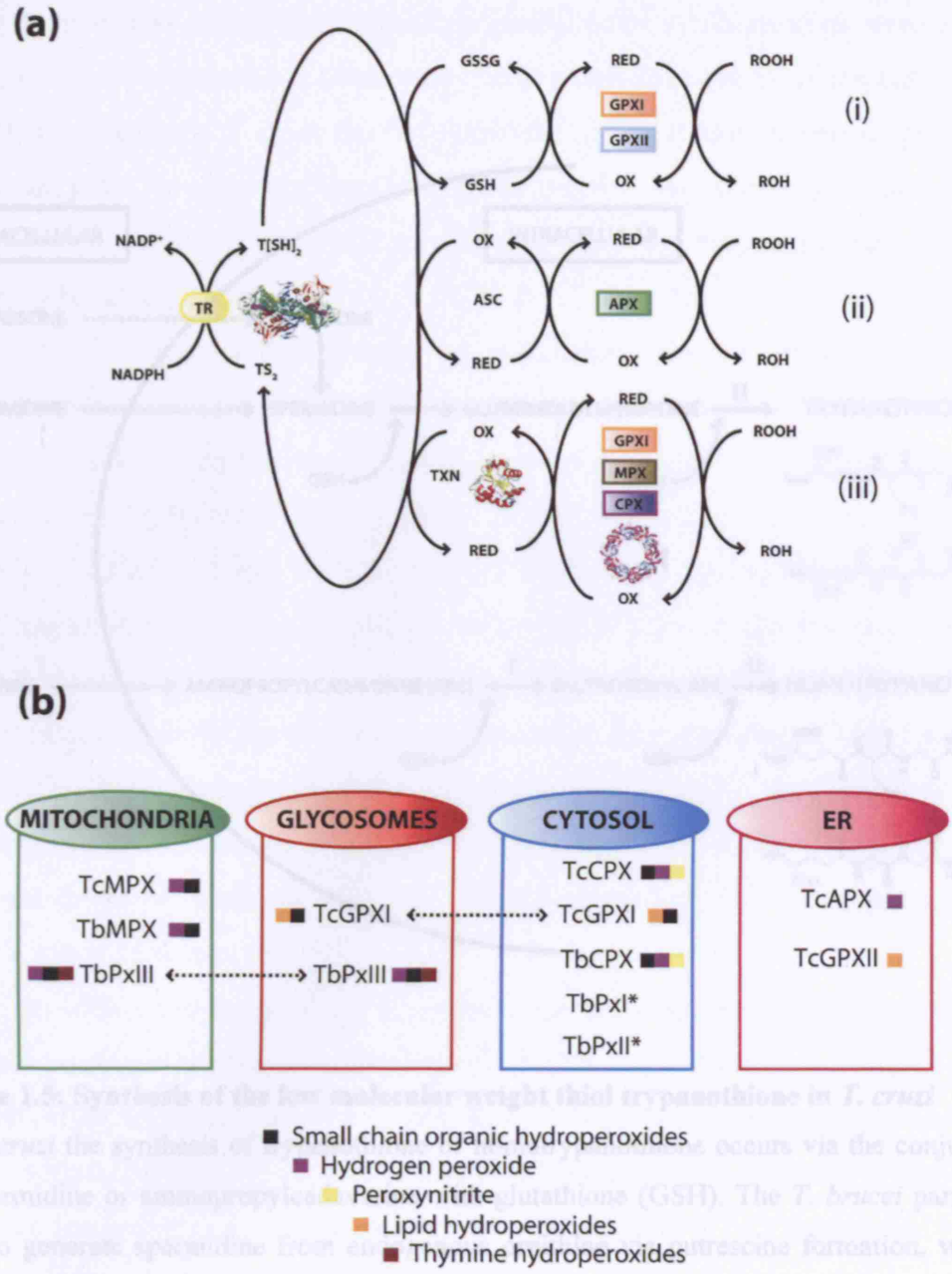


Figure 1.4: The trypanosomal oxidative defence mechanism

In the oxidative defence system of *T. cruzi* (a), reduction of the hydroperoxides occurs via the transfer of reducing equivalents from NADPH and dihydrotrypanothione ($T[SH]_2$) to either: (i) glutathione (GSH) and glutathione-dependent peroxidase (GPX) enzymes; (ii) ascorbate (ASC) and ascorbate-dependent peroxidase(APX); (iii) tryparedoxin (TXN) and cytosolic/mitochondrial peroxiredoxin (CPX/MPX) or GPX enzymes. In *T. brucei* APX is absent and the activity of the glutathione-dependent type peroxidases (TbPxI, TbPxII, TbPxIII) is linked only to tryparedoxin. Crystal structures of *T. cruzi* oxidised trypanothione reductase (TR) in complex with trypanothione (PDB: 1BZL), *T. brucei* TXN (PDB: 1O73) and *T. cruzi* CPX (PDB: 1UUL) are illustrated. The sub-cellular localisations and substrate specificities of the various trypanosomal peroxidase enzymes are shown (b).

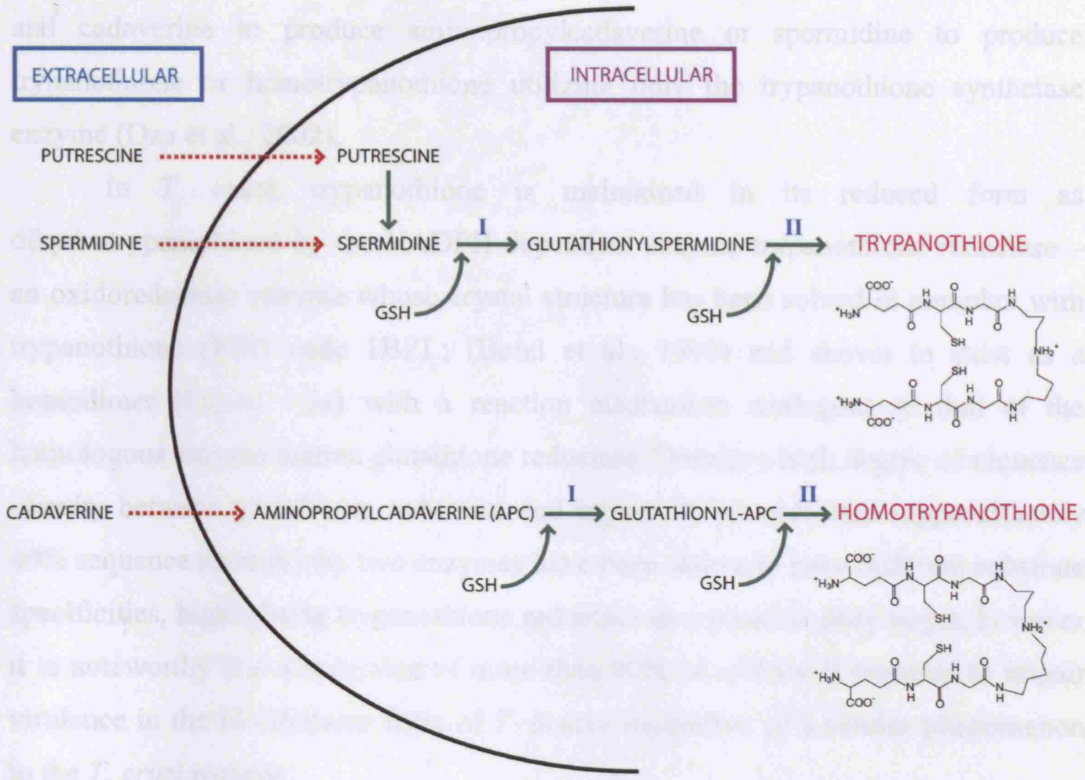


Figure 1.5: Synthesis of the low molecular weight thiol trypanothione in *T. cruzi*

In *T. cruzi* the synthesis of trypanothione or homotrypanothione occurs via the conjugation of spermidine or aminopropylcadaverine with glutathione (GSH). The *T. brucei* parasite is able to generate spermidine from endogenous ornithine via putrescine formation, whereas the *T. cruzi* parasite is unable to generate its own polyamines and thus transports exogenous putrescine, spermidine and cadaverine into the cell (dashed red arrows). In both *T. cruzi* and *T. brucei* parasites, the two intermediate steps (reactions I and II) in the formation of trypanothione are carried out by trypanothione synthetase instead of two separate enzymes, glutathionylspermidine synthetase and trypanothione synthetase as found in insect trypanosome *Crithidia fasciculata*. (Adapted from Oza et al., 2002).

biosynthesis differs in that polyamines are unable to be synthesised *de novo* and the parasite is unable to synthesis putrescine due to a deficiency in ornithine carboxylase activity. Consequently *T. cruzi* has developed the ability to use exogenous putrescine and cadaverine to produce aminopropylcadaverine or spermidine to produce trypanothione or homotrypanothione utilizing only the trypanothione synthetase enzyme (Oza et al., 2002).

In *T. cruzi*, trypanothione is maintained in its reduced form as dihydrotrypanothione by the NADPH-dependent enzyme trypanothione reductase – an oxidoreductase enzyme whose crystal structure has been solved in complex with trypanothione (PDB code 1BZL; (Bond et al., 1999) and shown to exist as a homodimer (Figure 1.6a) with a reaction mechanism analogous to that of the homologous enzyme human glutathione reductase. Despite a high degree of sequence identity between glutathione reductase and trypanothione reductase (approximately 40% sequence identity) the two enzymes have been shown to have different substrate specificities, highlighting trypanothione reductase as a possible drug target, however it is noteworthy that a reduction of more than 90% of activity is required to impair virulence in the bloodstream form of *T. brucei* suggestive of a similar phenomenon in the *T. cruzi* parasite.

In addition to the presence of classical thioredoxins (active site W-C-G-P-C), trypanosomes also contain an approximately 5 kDa larger thioredoxin-related dithiol protein called tryparedoxin (active site W-C-P-P-C), which is highly abundant in trypanosomes, particularly exhibited in *T. cruzi* and the insect-infecting trypanosome *Crithidia fasciculata* where tryparedoxin comprises >3-5 % of total soluble cellular protein. Tryparedoxin facilitates the reduction of hydroperoxides as a redox shuttle in the transfer of reducing equivalents from trypanothione to peroxiredoxin enzymes (Gommel et al., 1997) via a ping pong mechanism (typical of oxidoreductases) where the rate limiting step is the interaction with trypanothione. Tryparedoxin is shown to be essential as an electron donor: to maintain reduced forms of peroxidase enzymes in order to prevent sustained damage of cellular targets by hydroperoxides, to ribonucleotide reductase for the repair and synthesis of DNA (an analogous function to mammalian thioredoxin) and to universal minicircle sequence-binding protein (UMSBP) for the replication of the kinetoplast (Comini et al., 2007).

The reduced form of tryparedoxin has been shown to be essential in

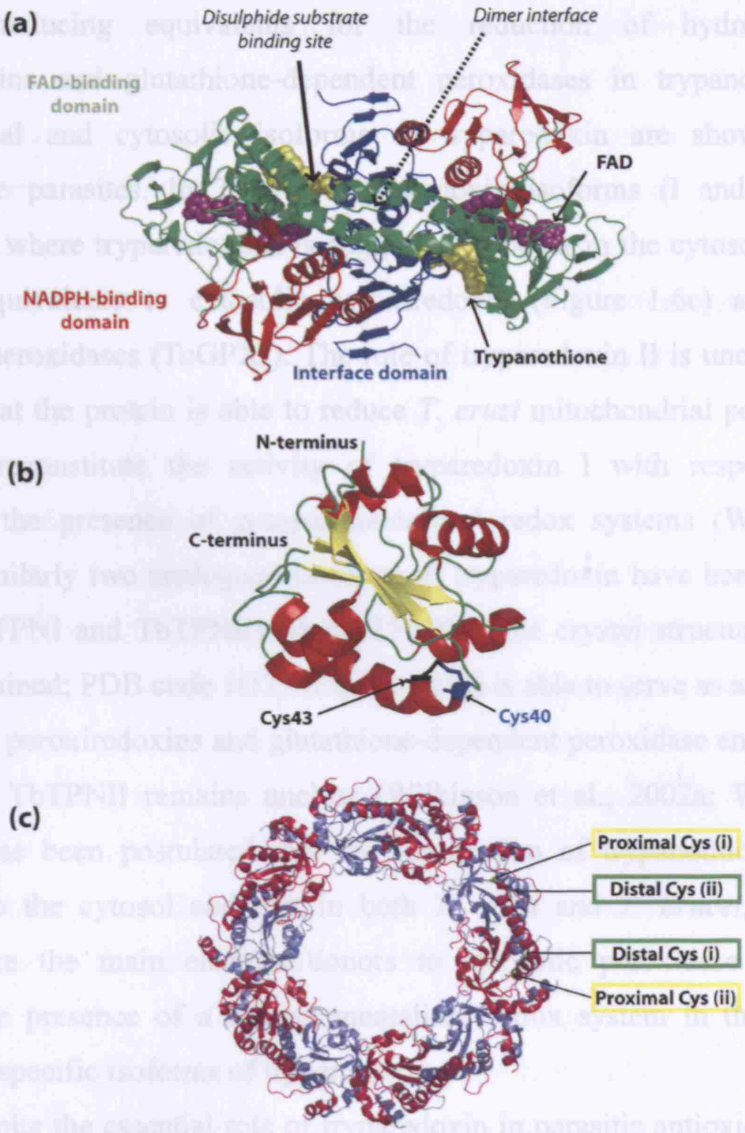


Figure 1.6: Crystal structures of components of the trypanosomal antioxidant system

Crystal structures of (a) *T. cruzi* oxidised trypanothione reductase (TR) in complex with trypanothione (PDB 1BZL), (b) *T. brucei* tryparedoxin (TXN) (PDB 1O73) and (c) *T. cruzi* cytosolic peroxiredoxin (CPX) (PDB 1UUL) are depicted as ribbon diagrams. In addition to the substrate-binding site, the various domains (NADPH-binding (red); FAD-binding (green); interface (red)), flavin adenin dinucleotide (FAD) molecules (purple) and trypanothione molecules (yellow) of homodimeric structure of TR are highlighted. The structure of *T. brucei* TXN shows the thioredoxin fold where catalytic cysteine residues (Cys40 (blue) and Cys43 (black)) are highlighted. The structure of *T. cruzi* CPX shows pentameric organisation of dimers (subunits of each dimer are shown in light blue (i) and pink (ii)) forming the classical decameric ring structure of typical 2-Cys peroxiredoxins. The proximal (yellow) and distal (green) redox cysteine residues contributed of each subunit in a dimer is highlighted.

providing reducing equivalents for the reduction of hydroperoxides via peroxiredoxins and glutathione-dependent peroxidases in trypanosoma. Various mitochondrial and cytosolic isoforms of tryparedoxin are shown to exist in trypanosome parasites. In *T. cruzi* tryparedoxin isoforms (I and II) have been established, where tryparedoxin I is suggested to occur in the cytosol and provide a reducing equivalents to cytosolic peroxiredoxin (Figure 1.6c) and glutathione-dependent peroxidases (TcGPXI). The role of tryparedoxin II is unclear, however it is known that the protein is able to reduce *T. cruzi* mitochondrial peroxiredoxin but unable to reconstitute the activity of tryparedoxin I with respect to TcGPXI suggesting the presence of compartmentalised redox systems (Wilkinson et al., 2000b). Similarly two analogous isoforms of tryparedoxin have been reported in *T. brucei* (TbTPNI and TbTPNII) where TbTPNI (the crystal structure of which has been determined; PDB code 1O73 (Figure 1.6b)) is able to serve as an electron donor to cytosolic peroxiredoxins and glutathione-dependent peroxidase enzymes, however the role of TbTPNII remains unclear (Wilkinson et al., 2002a; Wilkinson et al., 2003). It has been postulated that the localisation of trypanothione reductase is restricted to the cytosol and that in both *T. cruzi* and *T. brucei*, tryparedoxin I isoforms are the main electron donors to cytosolic peroxidase enzymes. This suggests the presence of a compartmentalised redox system in the mitochondria, which uses specific isoforms of tryparedoxin.

Despite the essential role of tryparedoxin in parasitic antioxidant defence, its role as a suitable drug target is questionable. This is particularly highlighted in *T. brucei* where studies have revealed that although cytosolic tryparedoxin activity is essential to parasite development against oxidative stress, tryparedoxin activity would have to be lowered to less than 5% in order to affect the virulence of the parasite (Wilkinson et al., 2000b). Additionally tryparedoxin, similar to all thioredoxin-like proteins, lacks a classical hydrophobic pocket making the design of inhibitors particularly challenging (Comini et al., 2007).

1.7 Peroxiredoxin (Prx)

In trypanosoma there are three classes of peroxidases that are dependent on trypanothione: 2-Cys peroxiredoxins, cysteine-homologues of the classic mammalian

selenocysteine-containing glutathione-dependent peroxidases and ascorbate-dependent peroxidase.

The ubiquitous peroxiredoxin (Prx) family are important in protection against reactive oxygen species in a wide variety of organisms, and are able to reduce a range of hydroperoxides including hydrogen peroxide and peroxynitrite species. In addition to peroxidase activity, the function of peroxiredoxins extends to involvement in cell proliferation, apoptosis and tumour suppressors.

1.7.1 Peroxiredoxin classification

In mammalian systems thioredoxin and glutaredoxin have been shown to serve as electron donors to peroxiredoxin isoforms, of which six members (Prx I-VI) have been reported with a range of localisations including peroxisomal, mitochondrial and cytosolic. A single gene typically encodes for the mitochondrial peroxiredoxin enzymes, whereas cytosolic isoforms tend to be encoded by clusters of genes at a chromosomal locus. Mammalian peroxiredoxins, except peroxiredoxin V, are homodimeric of subunit size 20-30 kDa and rely on either one or two conserved redox active cysteine residues for catalysis. The mammalian 2-Cys Prx enzymes contain a conserved redox active catalytic cysteine in an N-terminal V-C-P amino acid sequence motif, which is repeated near the C-terminus (in mitochondrial isoforms the C-terminus cysteine is contained in a conserved I-P-C motif instead of the V-C-P motif). In the 1-Cys peroxiredoxins only an N-terminal cysteine is conserved.

The peroxiredoxin family can be divided into three classes: typical 2-Cys peroxiredoxins; atypical 2-Cys peroxiredoxins and 1-Cys peroxiredoxins. The classification of the sub-classes is according to the number of redox active cysteine residues present (1-Cys or 2-Cys) with the further division of the 2-Cys class into 'typical' or 'atypical' (Figure 1.7).

The basic peroxidase mechanism is common to all peroxiredoxins whereby the N-terminal cysteine (peroxidatic cysteine, C_P) is activated by coordination with conserved arginine and threonine residues, facilitating the dissociation of the thiol group allowing attack by the hydroperoxide substrate and producing a cysteine sulfenic acid derivative. In the second stage of the reaction of typical 2-Cys

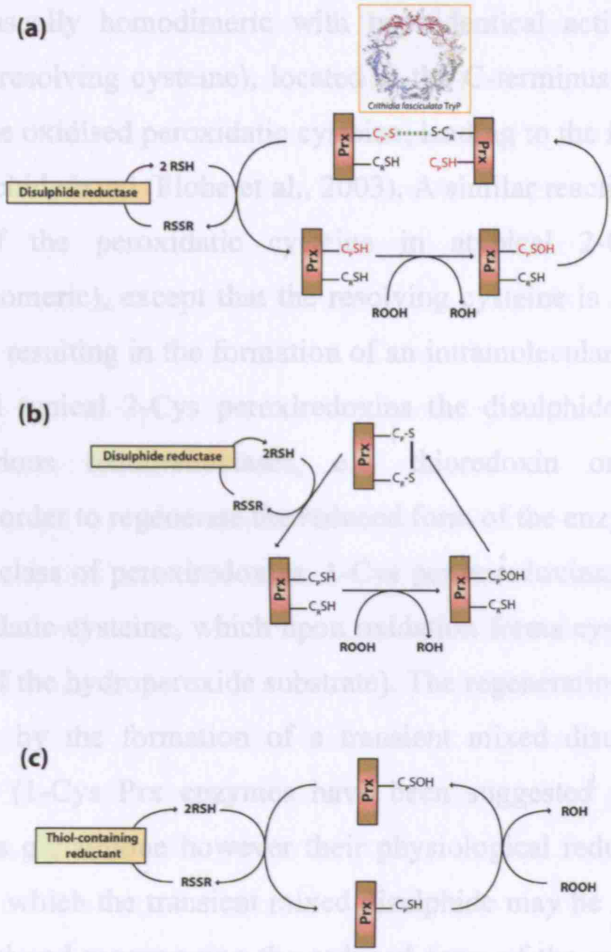


Figure 1.7: Catalytic mechanisms of the peroxiredoxin family

The mechanisms of the typical 2-Cys (a), atypical 2-Cys (b) and 1-Cys (c) classes of the peroxiredoxin enzymes (Prx) are shown. The first stage, universal to each class, is the reduction of the hydroperoxide substrate (ROOH) generating the sulphenic acid derivative (-SOH) of the peroxidatic cysteine (Cp). Following reduction of the hydroperoxide, the peroxidatic cysteine of typical and atypical 2-Cys Prx enzymes form an intermolecular (typical 2-Cys Prx) or intramolecular (atypical 2-Cys Prx) disulphide bond with a second resolving cysteine (C_R) which is subsequently reduced by a dithiol (RSH) (e.g. thioredoxin) which itself becomes oxidised (RSSR). Intermolecular disulphide bond formation in dimeric typical 2-Cys Prxs occurs between the N-terminal peroxidatic cysteine of one subunit and C-terminal resolving cysteine of a neighbouring cysteine in a head-to-tail manner (a), whereas monomeric typical 2-Cys Prxs contain a resolving cysteine in the same polypeptide chain (b). Regeneration of the reduced form of the oxidised reductant (RSSR) occurs via the action of flavoprotein disulphide reductase enzymes. Reduction of 1-Cys Prxs containing only one conserved cysteine (c), suggested to be preferentially performed by monothiols (e.g. glutathione) however the physiological reductants of these enzymes is equivocal. (Adapted from Wood et al., 2003).

peroxiredoxins (usually homodimeric with two identical active sites), a second cysteine residue (resolving cysteine), located at the C-terminus of another inverted subunit, attacks the oxidised peroxidatic cysteine, leading to the formation of a stable intersubunit disulphide bond (Flohe et al., 2003). A similar reaction occurs following the oxidation of the peroxidatic cysteine in atypical 2-Cys peroxiredoxins (functionally monomeric), except that the resolving cysteine is located on the same polypeptide chain resulting in the formation of an intramolecular disulphide bond. In both atypical and typical 2-Cys peroxiredoxins the disulphide bond can then be reduced by various oxidoreductases, e.g. thioredoxin or tryparedoxin (in trypanosomes) in order to regenerate the reduced form of the enzyme.

The third class of peroxiredoxins, 1-Cys peroxiredoxins, relies solely on the conserved peroxidatic cysteine, which upon oxidation forms cysteine sulphenic acid (from reduction of the hydroperoxide substrate). The regeneration of the ground state enzyme proceeds by the formation of a transient mixed disulphide with a thiol containing donor (1-Cys Prx enzymes have been suggested to prefer monothiol reductants such as glutathione however their physiological reductants are yet to be established), after which the transient mixed disulphide may be further reduced by a second donor thiol and regenerating the reduced form of the enzyme (Wood et al., 2003).

The mammalian peroxiredoxin enzymes are typical 2-Cys peroxiredoxins except for PrxV and PrxVI which are monomeric atypical 2-Cys and 1-Cys peroxiredoxins respectively (Wood et al., 2003; Declercq et al., 2001; Seo et al., 2000). A more detailed examination of the catalytic mechanism of typical 2-Cys peroxiredoxin enzymes has revealed the presence of conserved proline, arginine and threonine residues in close proximity to the peroxidatic cysteine. Both the arginine and threonine residues are shown to be vital for catalysis compared to the proline residue, which is suggested to play a role in a shielding effect of the sulfenic acid derivative in order to limit the access of the hydroperoxide substrate in causing further oxidation. In the case of 1-Cys mammalian peroxiredoxin enzymes, the sulfenic acid form of the enzyme is a stable intermediate, due to prevention of overoxidation by a nearby histidine residue. The 2-Cys peroxiredoxin enzymes are suggested to undergo local unfolding and conformational changes to expose the otherwise buried sulfenic acid group and facilitate the formation of a disulphide bond that is accessible to the suitable reductant.

Some typical 2-Cys peroxiredoxins have been found to undergo redox-sensitive oligomerisation e.g. bacteria, human and rat PrxII and *C. fasciculata* tryparedoxin peroxidase, where the peroxidatic cysteine is considered a molecular switch for this process. The reduced or overoxidised states have been shown to favour the formation of a pentameric configuration of dimers, associated with two sequence regions (designated A and B) found to be conserved only in typical 2-Cys Prx enzymes. Region A is part of the active site loop motif and forms a complementary surface to region B of the neighbouring dimer. In the reduced form, the peroxidatic cysteine located in the active site pocket is positioned against region A which is supported against region B. Upon reduction of the hydroperoxide substrate and formation of the sulfenic derivative where the sulphur and oxygen atom become buried, a conformational change occurs making the peroxidatic cysteine solvent exposed facilitating the formation of the disulphide intermediate. The crystal structure of human PrxII reveals the presence of a sulfinic group ($-SO_2H$), which traps the protein into a decamer and is thus unable to form a disulphide bridge. Overoxidation to sulfonic acid ($-SO_3H$) may lead to even further oligomerisation. This peroxiredoxin decameric structure is similarly observed in the reduced peroxiredoxin form, where the peroxidatic cysteine-containing loop region holds the decamer together, thus conformational changes result in instability and weakening of the decamer as more active sites form disulphide bonds, driving the formation of disulphide bond-associated dimeric structures (Wood et al., 2003).

1.7.2 Peroxiredoxins in trypanosoma

Trypanosomal peroxiredoxins belong to the class of typical 2-Cys peroxiredoxin enzymes with mitochondrial and cytosolic isoforms. As previously mentioned, *T. brucei* trypanothione reductase is suggested to be restricted to the cytosol, therefore the reduction of trypanosomal mitochondrial peroxiredoxin isoforms is suggested to be either maintained by a complete redox system in the mitochondria, or that the redox pathway is partitioned between the cytosol and mitochondria with the involvement of mitochondrial transport mechanisms (Kelly, 2003).

Similar to their mammalian counterparts the trypanosomal peroxiredoxin enzymes contain the conserved N- and C-terminus VCP motifs (IPC motif in mitochondrial isoforms) (Castro and Tomas, 2008). The crystal structures of various peroxiredoxins have been determined, including the cytosolic typical 2-Cys tryparedoxin peroxidase (TryP) from *Crithidia fasciculata* (PDB code 1E2Y; (Alphey et al., 2000) and *T. cruzi* cytosolic tryparedoxin peroxidase (PDB code 1UUL; (Pineyro et al., 2005) shown in Figure 1.6c. The trypanosomal peroxiredoxins exist in high abundance in these parasites (*i.e.* >5 % of cellular protein in *C. fasciculata*) and have a broad substrate specificity including the detoxification of hydrogen peroxide and organic hydroperoxides where cytosolic peroxiredoxins are also able to detoxify peroxynitrite, which is not formed in trypanosomes despite evidence for the presence of its precursors, superoxide anions and nitric oxide (Castro and Tomas, 2008). The *C. fasciculata* peroxiredoxin (CfTryP) has also been shown to reduce lipid hydroperoxides and similar to the typical 2-Cys Prx of human PrxII, is able to form a decamer of five dimers. The cytosolic *C. fasciculata* peroxiredoxin isoform is considered to be ideally localised for the detoxification of peroxides from the host immune system, whereas mitochondrial isoforms are specifically concerned with the detoxification of hydrogen peroxide formed as a bi-product of the respiratory system (Flohe et al., 2003). Studies of *Leishmania*, *T. cruzi*, *T. brucei* and *C. fasciculata* peroxiredoxins have shown decreased catalytic efficiency with increasing hydroperoxide concentrations – a phenomenon that can be linked to oligomerisation of the enzymes in conditions of overoxidation, similar to that observed in mammalian species.

In most eukaryotic cells peroxiredoxins function as regulators of hydrogen peroxide mediated cell signalling pathways, however there is uncertainty as to whether trypanosomal peroxiredoxins play a similar role, particularly highlighted by the fact that molecules such as sulfiredoxin (that are able to reduce overoxidised cysteine groups) are absent in the parasites. Therefore, the major function of the trypanosomal 2-Cys peroxiredoxin enzymes is considered to be in oxidative defence (Flohe et al., 2003; Castro and Tomas, 2008) where parasitic drug resistance and an increase in infectivity has been found to be associated with an upregulation in cytosolic and mitochondrial peroxiredoxins, suggesting these enzymes to be possible drug targets.

In *T. cruzi* the presence of trypanothione-dependent typical 2-Cys cytosolic and mitochondrial peroxiredoxin enzymes have been established where cytosolic isoforms are suggested to perform as a general peroxide scavenger and mitochondrial isoforms plays a protective role in the kinetoplast (Tetaud et al., 2001; Wilkinson et al., 2003; Wilkinson et al., 2000b). Both enzymes are able to reduce small chain organic hydroperoxides and hydrogen peroxide in protection of the parasite from endogenous and exogenous peroxide species, and utilise tryparedoxin in the regeneration of the ground state enzyme.

Similarly in *T. brucei* the cytosolic and mitochondrial isoforms are also able to reduce a variety of substrates including hydrogen peroxide. The cytosolic peroxiredoxin in *T. brucei* has been shown to be dependent on tryparedoxin and essential to parasite survival, however the mitochondrial isoform is considered to be non essential in the bloodstream form of the parasite and possibly compensated by the action of other peroxidases such as glutathione-dependent peroxidase enzymes (Wilkinson et al., 2003; Comini et al., 2007; Tetaud et al., 2001).

1.8 Ascorbate-dependent peroxidase (APX)

The parasite *T. cruzi* has been found to contain a plant-like haemoperoxidase, ascorbate peroxidase (TcAPX), however a homologous enzyme in *T. brucei* has not been found (Logan et al., 2007). Similar to the 2-Cys peroxiredoxins (reduces hydrogen peroxide and small chain organic hydroperoxide) and glutathione-dependent peroxidase (reduces fatty acid and phospholipid hydroperoxides), the activity of *T. cruzi* ascorbate-dependent peroxidase is dependent on the thiol trypanothione, which provides reducing equivalents to ascorbate peroxidase via ascorbate for the reduction of hydrogen peroxide generating the oxidised form of ascorbate, dehydroascorbate (DHA).

Ascorbic acid (the L-enantiomer is known as Vitamin C) is a cofactor for various enzymes involved in the absorption of iron, synthesis of collagen and neurotransmitters. In plants, ascorbic acid has been found to play a variety of roles including a role in collagen formation and iron absorption; however its main role is in the detoxification of reactive oxygen species. Ascorbate is a powerful reducing agent and reduces free radicals and some non-radical reactive species, as well as

providing reducing equivalents to ascorbate-dependent peroxidases (APX). Both *T. cruzi* and *T. brucei* have been shown to be able to synthesise vitamin C, where the former is able to produce ascorbate by *de novo* synthesis and the latter is able to additionally uptake the vitamin from its environment during the mammalian stages of infection. The pathway for the biosynthesis of Vitamin C in *T. brucei* and *T. cruzi* has been characterised and is shown to occur in the peroxisome-like organelle, the glycosome and involves the FAD-dependent family of oxidoreductases that convert aldonolactone to ascorbate (Wilkinson et al., 2005).

The APX family are class I haem-containing peroxidase enzymes which are known to detoxify hydrogen peroxides in plants and are monomeric with a molecular size of approximately 30 kDa. Although absent in *T. brucei*, both *Leishmania* and *T. cruzi* have both been found to contain ascorbate-dependent peroxidase, with sequence similarities to plant APXs in the range of 30-35% identity. The absence of APX in *T. brucei* is suggested to be due to the difference in infectivity between the African trypanosome, *T. cruzi* and *Leishmania* – the latter two parasites have intracellular mammalian stages, whereas *T. brucei* is solely extracellular, thus it is speculated that APX functions specifically in those parasites exposed to reactive oxygen species upon invasion of host macrophages.

Trypanosomal APX enzymes differ from their plant counterparts in that an arginine residue (Arg172 of soyabean cytosolic APX; accession number T07056) essential for binding and oxidation of ascorbate is absent, which is reflected in the low utilisation of ascorbate by these parasites. This difference suggests an alternative mechanism of ascorbate binding in trypanosomes. The catalytic mechanism of trypanosomal APX is speculated to be similar to that of plants where an extremely rapid second order reaction of ferric haeme with hydrogen peroxide occurs, thus resulting in the reduction of the hydroperoxide species, followed by the formation of an intermediate compound. The intermediate compound is reduced in two sequential electron transfer reactions by ascorbate, after which ferric haeme is formed and the ground state of the enzyme is regenerated. The catalytic reaction results in two monodehydroascorbate radicals, which in kinetoplastids can be reduced to ascorbate non-enzymatically via the spontaneous reduction of dehydroascorbate by trypanothione.

Similar to the *T. cruzi* glutathione-dependent peroxidase enzyme II (TcGPXII), the ascorbate-dependent peroxidase enzyme is localised in the

endoplasmic reticulum (Figure 1.4b), where both enzymes are suggested to play complementary roles in antioxidant defence (TcGPXII is able to metabolise fatty acid and phospholipid hydroperoxides and *T. cruzi* APX is able to metabolise hydrogen peroxide). The concentration of reduced ascorbate is maintained in the endoplasmic reticulum of *T. cruzi* via the transport of dehydroascorbate and subsequent conversion to the reduced form.

The *T. cruzi* APX enzyme is the first ascorbate-dependent haemoperoxidase to be found in a non-photosynthetic organism, which suggests that this enzyme arose from early evolution of the trypanosomatid lineage perhaps by an engulfment and gene transfer event. This is exhibited through phylogenetic analyses showing how kinetoplastida and euglenida together form the phylum Euglenozoa. Euglenoid algae contain plastids and studies have proposed that both the kinetoplastids and euglenoids acquired plastids by endosymbiosysis prior to divergence, however the kinetoplastids have lost the organelle but retained plant-like genes. The presence of plant-like genes in trypanosome is further supported by the similarity of other *T. cruzi* enzymes to their plant counterparts, particularly highlighted by *T. cruzi* GPXI (approximately 50 % sequence identity to higher plants). The absence of the plant-like APX from mammalian hosts thus provides a possible drug target (Wilkinson et al., 2002c; Hannaert et al., 2003).

1.9 Glutathione-dependent peroxidase enzymes (GPX)

In the host mammalian organisms, the reduction of hydrogen peroxides or organic hydroperoxides may be carried out by the family of glutathione peroxidase (GPX) enzymes.

1.9.1 Mammalian GPX enzymes

The mammalian GPX family are divided into seven clades according to their substrate specificity, primary sequence and localisation: cytosolic GPX (cGPX or GPX1), gastrointestinal GPX (GPX2), plasma GPX (GPX3), phospholipid GPX (PHGPX or GPX4), epididymal GPX (GPX5) and olfactory epithelium GPX (GPX6)

and most recently cytosolic GPX (GPX7) (Scheerer et al., 2007). The GPX family is proposed to have arisen from a PHGPX ancestor via gene duplication and genome integration events resulting in intracellular and secreted forms of GPX enzymes. Crystal structures of various mammalian GPX enzymes have been determined including bovine GPX1 (PDB code 1GP1), human GPX1 (PDB code 2F8A), human GPX2 (PDB code 2HE3), human GPX3 (2R37), human GPX4 (2GS3; 2OBI), human GPX5 (PDB code 2IY3) and human GPX7 (2P31).

Of the seven GPX members in humans, five contain a characteristic selenocysteine catalytic amino acid residue (GPX1-4 and GPX6) whilst the remaining two members (GPX5 and GPX7) contain a catalytic cysteine. Selenocysteine is considered to be the 21st amino acid and is cotranslationally incorporated into the protein where it contributes to the conserved glutathione peroxidase catalytic triad consisting of tryptophan, glutamine and selenocysteine (or cysteine). The action of selenocysteine has been shown to be essential for the efficient activity of these GPX enzymes, where cysteine mutations are found to cause a reduction of catalytic efficiency. This is attributed to the fact that at physiological pH the selenol group is fully ionized as opposed to the cysteine thiol group that is only partially ionised under similar conditions (Herbette et al., 2007). Phylogenetic studies suggest that ancestral GPX enzymes were cysteine-containing and the presence of the selenocysteine residue in some glutathione-dependent peroxidase enzymes has arisen more recently in evolution (Toppo et al., 2008).

1.9.2 Catalytic mechanism of GPX enzymes

The mechanism of the selenocysteine-containing GPX enzymes follows that of a ping-pong mechanism, which is characteristic of oxidoreductase enzymes. The mechanism begins with the oxidation of dissociated selenol group by hydroperoxide resulting in a selenic derivative, which then can react with glutathione resulting in a transient selenodisulphide bridge (Figure 1.8a). The selenodisulphide bridge may then be reduced by a second glutathione molecule, releasing the reduced form of GPX and oxidised glutathione (Maiorino et al., 1998). The other catalytic triad residues, glutamine and tryptophan, which are within hydrogen bonding distance of the active selenocysteine, provide protons for the dissociation/activation of the

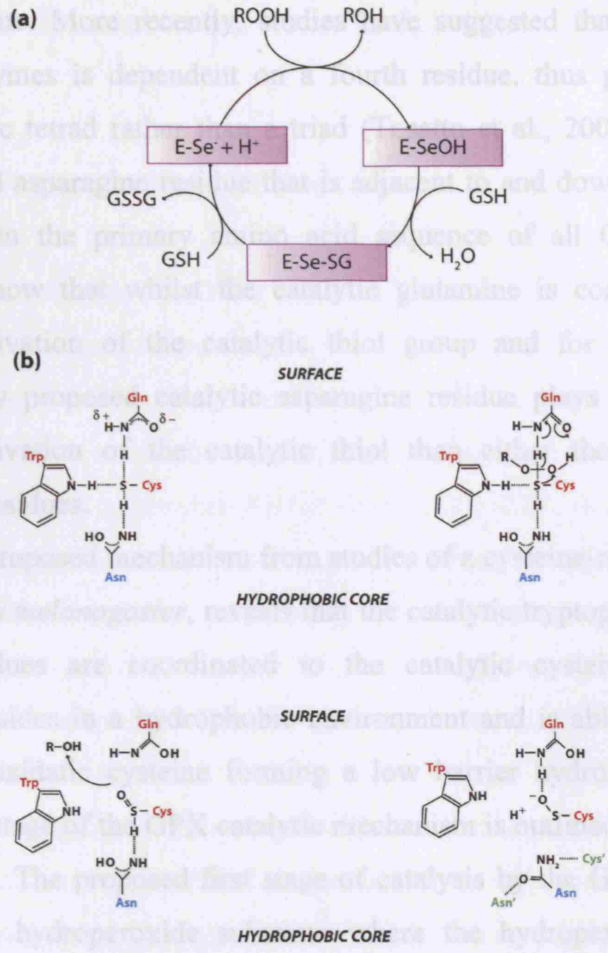


Figure 1.8: Catalytic mechanism of glutathione-dependent peroxidase (GPX) enzymes

The overall catalytic mechanism of phospholipid GPX (a) and closer examination of the recently proposed first stage of GPX catalysis utilising a catalytic tetrad, is illustrated (b). The GPX mechanism begins with the reduction of the hydroperoxide substrate (ROOH→ROH) generating oxidised GPX (selenenic acid derivative), which is reduced by two molecules of glutathione (GSH) via a transient selenodisulphide intermediate (E-Se-SG). In the reduction of hydrogen peroxide, an additional water molecule is formed during the initial hydroperoxide reduction. Closer examination of the first stage of catalysis (b) shows coordination/dissociation of the peroxidatic thiol group (selenocysteine or cysteine (red)) by conserved GPX catalytic residues glutamine (red), tryptophan (red) and recently proposed buried asparagine (cyan), with polarisation of the O-O bond of the substrate by glutamine facilitating reduction (ROH). The oxidised thiol of GPX is then structurally freed for interaction with thiol (or resolving cysteine) by alternative hydrogen bonding coordination with a conserved asparagine (Asn'; green) residue located in the sequence motif "NVA(S/T)" and cysteine residue (Cys'; green) located in sequence motif "FPCNQFG". (Adapted from Maiorino et al., 1998; Tosatto et al., 2008).

nucleophillic thiol and are involved in the orientation or polarisation of the hydroperoxide substrate. More recently, studies have suggested that the catalytic activity of GPX enzymes is dependent on a fourth residue, thus postulating the presence of a catalytic tetrad rather than a triad (Tosatto et al., 2008). The fourth residue is a conserved asparagine residue that is adjacent to and downstream of the catalytic tryptophan in the primary amino acid sequence of all GPX enzymes. Mutational studies show that whilst the catalytic glutamine is considered to be important in the activation of the catalytic thiol group and for with substrate interaction, the newly proposed catalytic asparagine residue plays an even more essential role in activation of the catalytic thiol than either the glutamine or tryptophan catalytic residues.

The recently proposed mechanism from studies of a cysteine-containing GPX enzyme in *Drosophila melanogaster*, reveals that the catalytic tryptophan, glutamine and asparagine residues are coordinated to the catalytic cysteine, where the asparagine residue resides in a hydrophobic environment and is able to attract the proton from the peroxidatic cysteine forming a low barrier hydrogen bond. The newly proposed first stage of the GPX catalytic mechanism is outlined in Figure 1.8b (Tosatto et al., 2008). The proposed first stage of catalysis by the GPX enzymes is the reduction of the hydroperoxide substrate where the hydroperoxide oxygen-oxygen bond is polarised by the surface exposed glutamine residue, where the relatively negative carbonyl group attracts the terminal oxygen proton and the amide group donates a proton to the proximal oxygen, thus generating an R-OH (alcohol) leaving group. Concurrently sulphur acts as a nucleophile attracted to the electron deficient oxygen, whereby sulfenic acid is formed (hydrogen bonded to the asparagine residue). Upon the dissociation and oxidation of the catalytic thiol, a series of proton shuffling events and changes in net hydrogen bridges occurs, which facilitates the oxidised thiol to be structurally released and move towards the resolving cysteine (in the case of GPX-type thioredoxin peroxidases) or to the reducing substrate (glutathione). The catalytic asparagine is able to free the oxidised peroxidatic thiol and participate in an alternative conformation by hydrogen bonding with another conserved asparagine (located in the conserved NVA(S/T) sequence motif upstream of the peroxidatic thiol) and cysteine (located in the conserved sequence motif FPCNQ). Despite the importance of the newly proposed catalytic asparagine-dependent proton shuffling for the first stages of the GPX catalytic

mechanism and in the reductive parts of the cycle, the overall rate-limiting step are reactions involving substrate interactions (Tosatto et al., 2008).

1.9.3 Substrate specificities of human GPX enzymes

The classic GPX enzymes are able to reduce hydrogen peroxide and organic hydroperoxides such as t-butyl hydroperoxide whereas the phospholipid GPX enzymes (PHGPX) are more efficient in catalysis of phospholipids and lipid peroxides. This difference in substrate specificity is thought to be due to the difference in oligomeric states between PHGPX and other GPX members. Human GPX 1-3 and GPX6 enzymes are homotetrameric containing dimer and tetramer interfaces that are absent in GPX4 and GPX7 (observed as two gap regions in multiple sequence alignments of GPX enzymes; Figure 1.9), thus these two enzymes (GPX4 and GPX7 are classed as PHGPX enzymes) are monomeric (except a PHGPX in *Populus trichocarpa* which has been shown to exist as a dimer despite the characteristic gaps in its amino acid sequence) with broad substrate specificities. The absence of the two regions that mediate oligomerisation in GPX4 and GPX7 leaves the active site more accessible to larger substrates in these enzymes, which may be the reason as to why both enzymes are able to metabolise complex lipid hydroperoxides. In addition to differences in substrate specificities in the GPX family, not all glutathione peroxidase enzymes utilise glutathione as an electron donor, as some GPX enzymes are able to accept reducing equivalents from alternative thiols such as thioredoxin or thioredoxin-related molecules. In fact, the residues implicated in the binding of glutathione are mainly observed in GPX1 (a lysine and four arginine residues surround the active site of GPX1), which has a high specificity for glutathione, whereas this specificity decreases in the order of GPX2, GPX3, GPX4 (in GPX4 mainly lysine residues are involved in glutathione specificity).

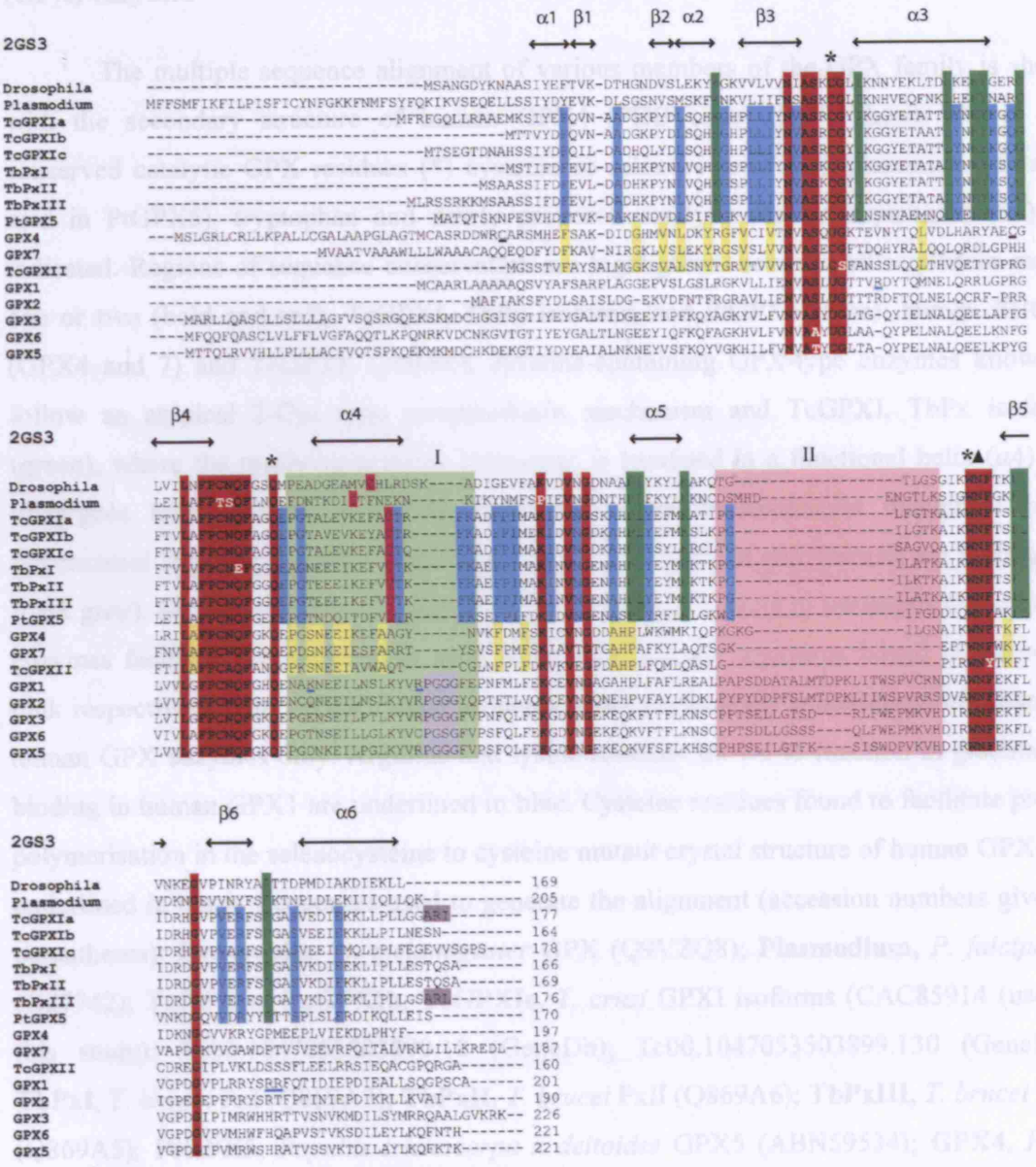


Figure 1.9: Multiple sequence alignment of glutathione-dependent type peroxidase (GPX) enzymes

(Figure legend overleaf)

Figure 1.9: Multiple sequence alignment of glutathione-dependent type peroxidase (GPX) enzymes

The multiple sequence alignment of various members of the GPX family is shown with the secondary structure of human GPX4 (PDB code 2GS3) shown above. The conserved catalytic GPX residues (*) cysteine or selenocysteine (U), glutamine (glutamic acid in PtGPX5), tryptophan and newly proposed catalytic residue asparagine (▲) are indicated. Regions of sequence conservation are highlighted: between all sequences except one or two (bold and red); TcGPXI, TbPx isoforms and PtGPX5 (blue); human PHGPX (GPX4 and 7) and TcGPXII (yellow); cysteine-containing GPX-type enzymes known to follow an atypical 2-Cys type peroxiredoxin mechanism and TcGPXI, TbPx isoforms (green), where the resolving cysteine (magenta) is localised in a functional helix (α4) that undergoes local unwinding for the formation of an intramolecular disulphide bond. Glycosomal targeting signals ("ARI" and "ARL") in TcGPXIa and TbPxIII are highlighted (dark grey). Sequence deletion regions (I and II) otherwise present in tetrameric human GPX enzymes facilitate oligomerisation, with dimer and tetramer interfaces boxed in green and pink respectively. The "PGGG"-type sequence motif (light grey) is conserved in tetrameric human GPX enzymes only. Arginine and lysine residues known to function in glutathione-binding in human GPX1 are underlined in blue. Cysteine residues found to facilitate protein polymerisation in the selenocysteine to cysteine mutant crystal structure of human GPX4 are underlined in black. Sequences used to generate the alignment (accession numbers given in parentheses): **Drosophila**, *D. melanogaster* GPX (Q9VZQ8); **Plasmodium,** *P. falciparum* (Q27742); **TcGPXIa**, **TcGPXIb**, **TcGPXIc,** *T. cruzi* GPXI isoforms (CAC85914 (used in this study); Tc00.1047053503899.10 (GeneDb); Tc00.1047053503899.130 (GeneDb)); **TbPxI**, *T. brucei* PxI (Q869A7); **TbPxII**, *T. brucei* PxII (Q869A6); **TbPxIII**, *T. brucei* PxIII (Q869A5); **PtGPX5**, *Populus trichocarpa x deltoides* GPX5 (ABN59534); **GPX4**, *Homo sapiens* GPX4 (P36969); **GPX7** *Homo sapiens* GPX7 (Q96SL4); **TcGPXII,** *T. cruzi* GPXII (CAC85915); **GPX1,** *Homo sapiens* GPX1 (PO7203); **GPX2,** *Homo sapiens* GPX2 (P18283); **GPX3**, *Homo sapiens* GPX3 (P22352); **GPX6**, *Homo sapiens* GPX6 (P59796); **GPX5**, *Homo sapiens* GPX5 (O75715).

1.9.4 Cysteine-containing GPX enzymes

Some GPX enzymes contain an active site cysteine instead of the active site selenocysteine and occur in various organisms including mammals but predominantly in plants, yeast and trypanosoma. Most non-selenium GPX members belong to the PHGPX sequence family (except, for example, mammalian GPX5 and a nematode GPX which are closer sequence homologues to GPX3) however some cysteine-containing GPX enzymes such as those occurring in *T. cruzi* and plants have been shown to exhibit low PHGPX activity, compared to mammalian PHGPX.

Studies have indicated that the residues required for the binding of glutathione in PHGPX and PHGPX-like proteins are absent, and in fact GPX4 has been found to have a low affinity for glutathione, similar to other non-selenium GPX enzymes in plants and trypanosoma suggesting that some non-selenium GPX family members are able to use electron donors other than glutathione (Brigelius-Flohe et al., 1994). This has led to the proposal for a reclassification of non-selenium GPX enzymes as thiol peroxidases rather than glutathione-specific peroxidases, as a separate group from the classic GPX family. This is further supported by the evolutionary studies showing that non-selenium GPX enzymes cluster with mammalian PHGPX forming a separate clade from selenium-containing GPX members. The selenocysteine GPX family are shown to occur mainly in vertebrates, whilst cysteine-containing GPX enzymes are predominant in non-vertebrates with the majority sharing common structural features that suggest a thioredoxin preference rather than glutathione (Tosatto et al., 2008).

The GPX-type thioredoxin peroxidases are mainly monomeric containing an essential conserved second cysteine, analogous to the resolving cysteine in the atypical 2-Cys peroxiredoxins, located downstream of the peroxidatic cysteine in a helical region that undergoes a local unwinding to facilitate the formation of an intramolecular disulphide bond with the peroxidatic cysteine. Research on a cysteine-containing GPX homologue in *Drosophila melanogaster*, that lacks glutathione binding residues and shows only slight glutathione activity, has been used to provide a basis for the preferential use of thioredoxin by some GPX homologues, rather than glutathione as is typical for mammalian selenocysteine GPX enzymes (Maiorino et al., 2007).

The formation of an intramolecular disulphide bond between the peroxidatic and resolving cysteine residues in GPX-type thioredoxin peroxidases is suggested to be a requirement for thioredoxin specificity. Studies have proposed the following criteria to characterise the non-selenium GPX-type thioredoxin peroxidases: deletions of oligomerisation interfaces (these enzymes are predominantly monomeric); a flexible loop region which forms the dimer interface in mammalian tetrameric GPX enzymes and contains a helix; a resolving cysteine (the resolving cysteine is not required by typical mammalian selenocysteine GPX enzymes) located downstream of the peroxidatic cysteine in the flexible loop region that is able to form a disulphide bond with the peroxidatic cysteine. Sequence analyses have revealed that the majority of GPX enzymes found in higher plants, protozoa, bacteria and fungi satisfy these features and are thus likely to function as thioredoxin peroxidases (Maiorino et al., 2007) . This has been also exemplified in a cysteine-containing GPX enzyme in *Plasmodium falciparum* that exhibits low glutathione affinity, which is able to reduce hydrogen peroxides and organic hydroperoxides in a thioredoxin-dependent manner. This parasitic GPX enzyme uses a downstream second cysteine for the formation of a disulphide bridge, which can be reduced by thioredoxin following reduction of the hydroperoxide substrate.

Various isoforms of cysteine-containing GPX-like enzymes have been characterised (Figure 1.4b) in *T. cruzi* and *T. brucei* (*T. cruzi* GPXI and II; *T. brucei* PxI, PxII, PxIII), that are similar in sequence to the PHGPX family but show low PHGPX activity, poor affinity to glutathione and greater affinity for the thioredoxin-related tryparedoxin (Wilkinson and Kelly, 2003; Schlecker et al., 2005). Investigations into the catalytic mechanism of the tryparedoxin-dependent *T. brucei* PxIII, has shown this enzyme to employ an atypical 2-Cys peroxiredoxin type mechanism characterised by the formation of an intramolecular disulphide bond between the peroxidatic and resolving cysteines, following reduction of the hydroperoxide (Schlecker et al., 2007). In *T. brucei* PxIII the catalytic residue glutamine is shown to be essential to catalysis, whereas tryptophan is suggested to play a structural role.

The ability of some GPX enzymes to use thioredoxin (e.g. thioredoxin-dependent cysteine-containing GPX in *Plasmodium falciparum*, *Brassica napus* and *Saccharomyces cerevisiae*) and of some peroxiredoxin enzymes to use glutathione (peroxiredoxin enzymes typically utilise thioredoxin as a reductant) suggests

overlapping and complementary systems between GPX and peroxiredoxin enzymes (Herbette et al., 2007).

1.9.4.1 *T. cruzi* glutathione-dependent peroxidase I (TcGPXI)

In the parasite *T. cruzi* a genomic locus encoding a cluster of three near identical genes (95 % sequence identity) coding for glutathione-dependent type peroxidase enzymes has been identified. One of these genes has been well characterised and codes for an approximately 18 kDa protein called *T. cruzi* glutathione-dependent enzyme I (TcGPXI) (Schlecker et al., 2005; Wilkinson et al., 2000a).

Sequence analysis of TcGPXI shows a conserved catalytic triad consisting of cysteine, glutamine and tryptophan with characteristic deletion regions suggesting a monomeric nature of the enzyme, similar to that of the PHGPX family. The TcGPXI enzyme is found to lack known glutathione-binding residues similar to other PHGPX family members. Although the *T. cruzi* parasite is found to not contain significant levels of glutathione unlike the host mammalian cells, TcGPXI is able to use glutathione as a source of reducing equivalents, albeit with low efficiency compared to mammalian GPX enzymes. Trypanosomes have been found to lack glutathione reductase (present in mammalian counterparts) and therefore utilise trypanothione or a trypanothione-glutathione thioltransferase (*e.g.* Tc52 in *T. cruzi*) to recycle oxidised glutathione back into the reduced form. The TcGPXI enzyme has been shown to be able to reduce fatty acid and phospholipid hydroperoxides including short chain alkyl hydroperoxides (e.g. cumene hydroperoxide and t-butyl hydroperoxide), via the transfer of electrons from NADPH and trypanothione to either glutathione or thioredoxin-like tryparedoxin (TcTPNI). TcGPXI displays a ping-pong type mechanism with tryparedoxin and has a higher affinity for tryparedoxin than glutathione (activity rate 8-15 fold higher than glutathione-dependent catalysis) (Wilkinson and Kelly, 2003).

The primary sequence of TcGPXI contains the sequence motif "ARI" ("ARL" in TbPxIII) at the C-terminus (Figure 1.9) corresponding to a glycosome targeting signal which is coherent with studies showing that as well as being present in the cytosol, TcGPXI is localised in the glycosome (the other two TcGPXI

isoforms are suggested to occur only in the cytosol). This may reflect the ability of the TcGPXI enzyme to use different electron donors (glutathione or tryparedoxin) according to subcellular localisation. The use of alternative electron donors according to subcellular localisation is also highlighted in the mammalian peroxiredoxin V enzyme, which utilises reducing equivalents from thioredoxin when localised in the cytosol and glutathione when localised in peroxisomes. The glutathione-dependent type peroxidase TcGPXI, is unable to metabolise hydrogen peroxide, thus it is likely that antioxidant defence is achieved by the complementary action of ascorbate-dependent peroxidase and peroxiredoxin enzymes against hydrogen peroxide, where glutathione-dependent peroxidase enzymes function against secondary products from lipid peroxidation.

1.9.4.2 *T. cruzi* glutathione-dependent peroxidase II (TcGPXII)

A second genomic locus in *T. cruzi* is found to contain a gene encoding a second glutathione-dependent peroxidase enzyme, *T. cruzi* GPXII (TcGPXII) of molecular weight 16 kDa (Wilkinson et al., 2002b). The TcGPXII enzyme, similar to TcGPXI, contains the conserved catalytic triad of cysteine, glutamine and tryptophan, deletion regions suggested to contribute to the monomeric nature of the protein and is found to be similar in sequence to the PHGPX family. Despite this similarity to the PHGPX family, TcGPXII is found to be more similar to a range of GPX enzymes from other organisms (31-37 % sequence identity) than to TcGPXI (sequence identity of 33%). TcGPXI shows 50 % and 72% sequence identity to plant and *T. brucei* GPX enzymes respectively, suggesting different evolutionary origins for TcGPXI and TcGPXII in the trypanosomal parasite. Unlike TcGPXI, the TcGPXII enzyme is able to only function in a glutathione-dependent manner displaying a ping-pong type mechanism, where oxidised glutathione is recycled into the reduced form via trypanothione. The enzyme is localised in the endoplasmic reticulum where the concentration of reduced glutathione is found to be greater than in the cytosol. TcGPXII catalyses the reduction of fatty acid and phospholipid hydroperoxides but is unable to reduce short chain alkyl hydroperoxides. Similar to TcGPXI, the TcGPXII enzyme is unable to reduce hydrogen peroxide but unlike TcGPXI is localised in the endoplasmic reticulum, thus complemented by the

function of ascorbate-dependent peroxidase and speculated to play a specific role in protection from oxidative damage of newly synthesised lipids (Wilkinson and Kelly, 2003).

1.9.4.3 *T. brucei* glutathione-dependent peroxidases

A genomic locus on chromosome 7 of approximately 4.4 kb in size contains a cluster of three approximately 95 % identical genes encoding isoforms of cysteine-homologues of glutathione-dependent type peroxidases in *T. brucei* (TbPxI, TbPxII, and TbPxIII). Overall the amino acid sequences (166, 169 and 176 amino acid residues for TbPxI, TbPxII, TbPxIII respectively) are near identical with most sequence variation occurring at the N- and C-termini (Hillebrand et al., 2003). Sequence analysis of the *T. brucei* peroxidase enzymes shows similarity to glutathione-dependent peroxidases of other organisms particularly to the plant GPX family, with identification of a conserved catalytic triad comprising cysteine, glutamine and tryptophan residues. The *T. brucei* peroxidase enzymes may be classified as part of the PHGPX family and exhibit characteristic sequence deletions associated with monomeric GPX proteins.

The largest and most well characterised enzyme of the *T. brucei* peroxidase isoforms is monomeric TbPxIII with a molecular weight of approximately 19 kDa. The *T. brucei* peroxidases are close homologues of TcGPXI and show high sequence identity to plant GPXs suggesting similar evolutionary events in both parasites. Whilst both TcGPXI and TbPxIII contain glycosomal targeting signals (sequence motifs "ARI" (TcGPXI) and "ARL" (TbPxIII)) the *T. brucei* enzyme has been shown to be localised in the mitochondria and cytosol, suggesting a suppression of the glycosomal targeting signal (Schlecker et al., 2005) although a recent review proposes that TbPxIII is localised in the glycosomes and mitochondria, whereas TbPxI and TbPxII reside in the cytosol (Castro and Tomas, 2008).

T. brucei peroxidase enzymes, like TcGPXI, lack known glutathione-binding residues and reduce hydroperoxide substrates by a ping-pong mechanism via reducing equivalents from *T. brucei* thioredoxin-like tryparedoxin (TbTPNI) and trypanothione, where the activity of TbPxIII is five fold greater with tryparedoxin than with glutathione. Unlike TcGPXI, *T. brucei* PxIII has a substrate preference for

hydrogen peroxide and additional specificity for substrates including t-butyl hydroperoxide, thymine hydroperoxide, linoleic acid hydroperoxide but poor substrate specificity for phosphatidylcholine hydroperoxide. The *T. brucei* GPX-type peroxidase enzymes have been established as being essential for parasite survival, thus it may be that essential mitochondrial GPX isoforms are able to compensate for the otherwise apparently non-essential mitochondrial 2-Cys peroxiredoxin enzyme in the parasite (Wilkinson et al., 2003).

1.9.4.3.1 Peroxiredoxin-type catalytic mechanism of *T. brucei* GPX-type peroxidase

The catalytic mechanism of TbPxIII has recently been elucidated by mutational and kinetic analyses. Although TbPxIII is classified as a cysteine-homologue of the classical glutathione peroxidases, the enzyme is shown to display a mechanism that resembles that of the atypical 2-Cys peroxiredoxin type mechanism, which is also observed in other GPX-type thioredoxin peroxidase enzymes. The TbPxIII enzyme contains a catalytic peroxidatic cysteine (Cys47), and additional conserved cysteine residues in the "FPCNQF" motif (Cys76) and in the F(Y)-A(V)-CT motif (Cys95), the latter of which is conserved only in peroxidases of yeast, plants, fungi and trypanosomes. In the TbPxIII enzyme, the first stage of catalysis is the reduction of the hydroperoxide substrate and subsequent oxidation of the peroxidatic cysteine (Cys47) to sulfenic acid. The second stage proceeds by the attack of Cys95 (resolving cysteine) on the oxidised peroxidatic cysteine, resulting in the formation of an intramolecular disulphide bond. The resolving cysteine, Cys95, is then attacked by the redox-active Cys40 of tryparedoxin resulting in a mixed disulphide and regeneration of the Cys47 thiol group. Following the formation of the mixed disulphide intermediate, a second redox-active cysteine of tryparedoxin (Cys43) attacks the nearby tryparedoxin cysteine (Cys40) participating in the mixed disulphide, to release oxidised tryparedoxin and regenerating the thiol group of Cys95 (Schlecker et al., 2007). In a GPX enzyme present in Chinese cabbage and homologous to TbPxIII, a three cysteine mechanism is proposed whereby the intramolecular disulphide bridge forms between the resolving cysteine and the second cysteine located in the "FPCNQF" motif. However, in TbPxIII a three

cysteine mechanism is shown to be unlikely and the analogous second cysteine in the *T. brucei* enzyme is suggested to play a structural or regulatory role in catalysis instead (Jung et al., 2002). The conserved glutamine of the TbPxIII catalytic triad is shown to be essential for catalysis whereas mutation of the conserved tryptophan allows catalytic activity to occur albeit at a reduced level compared to that of the wild type enzyme, suggestive of a structural role of tryptophan in catalysis (Schlecker et al., 2007).

The high sequence identity between TbPxIII and TcGPXI (76 % sequence identity) may allow inference of similar tertiary structures and catalytic mechanisms occurring in these two enzymes with possible differences in structural regions around the redox centre attributing to the differences in substrate specificities.

1.10 Structures of known homologues of trypanosomal GPX enzymes

The crystal structures of various glutathione peroxidase enzymes have been determined where GPX homologues of known structure may allow inference into the tertiary structure and possible catalytic mechanism of known trypanosomal GPX-type peroxidase enzymes. The closest *T. cruzi* GPXI and *T. brucei* peroxidase sequence homologues (sequence identity ranges 48-52 %) of determined structure are the reduced (PDB code: 2P5Q) and oxidised (PDB code: 2P5R) conformational forms of a plant thioredoxin peroxidase from *Populus trichocarpa x deltoids* (PtGPX5) shown in Figure 1.10. In contrast to this, the closest sequence homologue of known structure to *T. cruzi* GPXII is human GPX4 (PHGPX) where structures have been determined that contain mutations of the catalytic selenocysteine to either an inactive glycine (PDB code: 2GS3) or to a less catalytically efficient cysteine (PDB code: 2OBI) residue (Figure 1.11). More recently the crystal structure of an additional *T. cruzi* GPXII homologue, which is the cysteine-containing human GPX7 (similar to GPX4) enzyme, has been determined which exhibits the same sequence identity to *T. cruzi* GPXII, as does human GPX4 (sequence identity is 42 %).

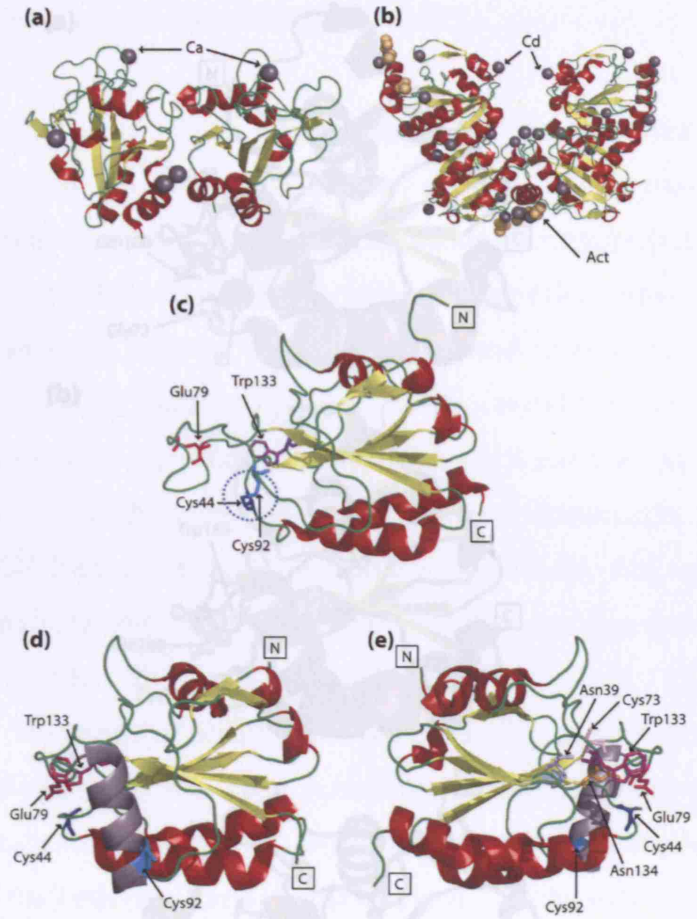


Figure 1.10: Crystal structures of the oxidised and reduced conformations of *Populus trichocarpa x deltoides* GPX5

The crystal structures of (a) oxidised (PDB: 2P5R) and (b) reduced (PDB: 2P5Q) PtGPX5 are depicted as ribbon diagrams. Calcium ions (Ca) from crystallisation solution are highlighted in the oxidised form (grey spheres); cadmium ions (Cd) and acetate ions (Act) from crystallisation solution are highlighted in the reduced form (grey and orange spheres respectively). The oxidised form (c) contains an intramolecular disulphide bond (dashed blue circle) between the peroxidatic cysteine (Cys44; blue) and resolving cysteine (Cys92; cyan) in addition to the conserved catalytic residues glutamic acid (Glu79; magenta) and tryptophan (Trp 133; purple). The reduced conformation (d) shows the resolving cysteine (Cys92; cyan) localised in a functional helix (grey), which unravels to facilitate the formation of the disulphide bond in the oxidised state. Additional residues (Asn39 (light blue); Cys73 (light pink)) analogous to residues suggested to participate in the mechanism for the first stage of GPX catalysis in the homologue *D. melanogaster* GPX-type enzyme are shown, including the conserved asparagine (Asn134; orange) contributing to the proposed GPX catalytic tetrad, are highlighted on the reduced conformation structure (e).

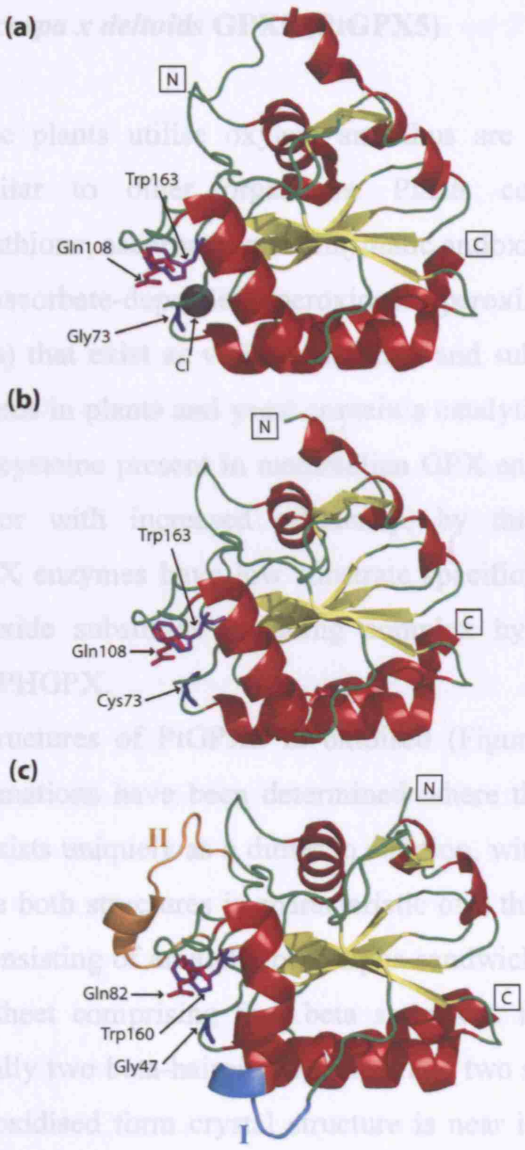


Figure 1.11: Crystal structures of human GPX4 and human cellular GPX1

The crystal structures of human GPX4 (a: PDB: 2GS3 and b: PDB: 2OBI) and human cellular GPX1 (c: PDB: 2F8A) are shown as ribbon diagrams, where a chloride ion (Cl) is depicted as a grey sphere (a). The crystal structures of human GPX4 have been determined where the catalytic selenocysteine is mutated to (a) glycine (Gly73; blue) and (b) cysteine (Cys73; blue) are indicated. Additional GPX catalytic residues, glutamine (Gln108; magenta) and tryptophan (Trp163; purple) are highlighted. The crystal structure of human cGPX1 (c) contains the conserved catalytic triad of selenocysteine (mutated to glycine (Gly47; blue)), glutamine (Gln82; magenta), tryptophan (Trp160; purple). Human cGPX1 contains loop regions I (cyan) and II (brown) that facilitate oligomerisation of some GPX enzymes. Region II is suggested to restrict the access of complex lipid hydroperoxides to the active site. The deletion of regions I and II accounts for the monomeric nature and wider substrate specificity of PHGPX isoforms.

1.10.1 *Populus trichocarpa x deltoids* GPX5 (PtGPX5)

Photosynthetic plants utilise oxygen and thus are subjected to reactive oxygen species similar to other organisms. Plants contain non-enzymatic antioxidants (e.g. glutathione, ascorbate) and enzymatic antioxidants (e.g. superoxide dismutase, catalase, ascorbate-dependent peroxidase, peroxiredoxins, glutathione-dependent peroxidases) that exist as various isoforms and subcellular localisations. Most GPX-type enzymes in plants and yeast contain a catalytic cysteine as opposed to the catalytic selenocysteine present in mammalian GPX enzymes and are able to be reduced solely, or with increased efficiency, by thioredoxin rather than glutathione. Plant GPX enzymes have low substrate specificity and reduce a wide variety of hydroperoxide substrates including complex hydroperoxides and are similar to monomeric PHGPX.

The crystal structures of PtGPX5 in oxidised (Figure 1.10a) and reduced (Figure 1.10b) conformations have been determined where the polypeptide of 170 amino acid residues exists uniquely as a dimer in solution, with equivalent subunits. The overall fold of the both structures is characteristic of a thioredoxin fold, typical of the GPX family, consisting of an alpha-beta-alpha sandwich. In the reduced form central twisted beta sheet comprising four beta strands is flanked by four alpha helices, with additionally two beta-hairpin structures and two smaller 3/10 helices at the N-terminus. The oxidised form crystal structure is near identical to that of the reduced conformation except for an alpha helical region containing the resolving cysteine in the reduced form (Figure 1.10d), which is unwound forming a large loop structure facilitating a conformational change for the formation of an intramolecular disulphide bond (Figure 1.10c) in the oxidised fold. In the reduced conformation, the side chain of the peroxidatic cysteine in the active site pocket faces the interior of the protein, surrounded by mainly non-polar groups (including catalytic residue Trp133) and hydrophilic catalytic residues Glu79 and Asn134. The peroxidatic cysteine is solvent accessible due to its localisation near the subunit surface, which then forms a disulphide bond with the resolving cysteine upon reduction of the hydroperoxide substrate. The crystal structure of oxidised PtGPX5 shows that one side of the disulphide bond faces the aromatic residue Phe95, while the other side is solvent exposed so that interaction with thioredoxin may occur.

1.10.1.1 The peroxiredoxin-like catalytic mechanism of *Populus trichocarpa x deltoids* GPX5 (PtGPX5)

The crystal structures of PtGPX5 reveals the presence of the conserved GPX catalytic triad (cysteine, glutamic acid, tryptophan) where glutamic acid replaces the catalytic glutamine found in mammalian GPX structures. This adaptation in the catalytic triad and unique non-covalent dimerisation, which occurs despite sequence deletions that otherwise mediate oligomerisation in other GPX enzymes, is suggested to be due to adaptations for substrate and thioredoxin interactions. Despite sequence similarities of higher plant GPX enzymes to their mammalian counterparts, which contain a selenocysteine and exist as tetramers (except PHGPX), many plant GPX enzymes possess a peroxidatic cysteine and a resolving cysteine suggesting that these enzymes participate in a mechanism that resembles that of the atypical 2-Cys peroxiredoxin. The formation of the intramolecular disulphide bond intermediate is considered to be an essential requirement for the ability of some GPX-type peroxidase enzymes to utilise thioredoxin, thus PtGPX5 has been shown to be a thioredoxin peroxidase although structurally characteristic of the GPX family.

From the structural determination of PtGPX5, the catalytic mechanism is proposed to begin by the reduction of the hydroperoxide substrate and generation of sulfenic acid derivative of the peroxidatic cysteine (Cys44), which rapidly reacts with the resolving cysteine (Cys92) via the unwinding of a helix (containing the resolving cysteine) and a significant conformational change. Figure 1.10d shows the reduced conformational tertiary structure of PtGPX5, where the conserved GPX catalytic triad and additional residues that are analogous to those residues involved in the recently proposed GPX mechanism utilising a catalytic tetrad are highlighted (Figure 1.10e). The resultant intramolecular disulphide bond formed in the oxidised conformation is subsequently reduced by interaction with thioredoxin, where a tryptophan residue at the active site of thioredoxin plays an important role in recognition. The proximal cysteine of thioredoxin is suggested to form an intermolecular disulphide bond with the PtGPX5 resolving cysteine following recognition of a proposed thioredoxin recognition motif EPGxx(D/E)xIxx(F/M)(V/A)CT(R/K)FK (where x denotes any amino acid) in PtGPX5, which is also found to be conserved in the majority of plant GPX enzymes (Koh et al., 2007). The second cysteine of thioredoxin is then able to attack the

intermolecular disulphide bond generating both the ground state PtGPX enzyme and oxidised thioredoxin containing an intramolecular disulphide bond, where one side of the thioredoxin disulphide is associated with charged groups (proposed site of interaction with GPX substrate) and is distinct from the other side of the thioredoxin disulphide bond which has a flat and hydrophobic surface (proposed site of interaction with thioredoxin reductase). The peroxiredoxin-like catalytic mechanism and redox-driven conformational changes exhibited in PtGPX5 are likely to occur in many higher plants that share biochemical functional and sequence similarities, as well as in other GPX-type thioredoxin peroxidases of other organisms.

1.10.2 Human phospholipid GPX4

In addition to the structure of PtGPX5, the crystal structure of the trypanosomal homologue human GPX4 (PHGPX) has been determined. Monomeric human GPX4 has been shown to exist as many isoforms in various subcellular localisations (cytosolic, mitochondrial, nuclear) and is uniquely able to reduce complex lipid hydroperoxides, as opposed to other human GPX members that are mostly tetrameric and have a less broader substrate specificity. Similar to other human GPX members, PHGPX contains an active site selenocysteine residue in addition to glutamine and tryptophan, which together comprise the conserved GPX catalytic triad. Two crystal structures of human GPX4 have been solved (Figure 1.11a and b), where selenocysteine has been mutated to glycine (catalytically inactive; PDB code 2GS3) and cysteine (reduced catalytic activity; PDB code 2OBI) at resolutions of 1.9 and 1.55 Å respectively. Despite the differences in catalytic activity between both proteins, their overall structures have significantly similar conformations (Scheerer et al., 2007).

The mutation of the active site selenocysteine to cysteine in human GPX4 has revealed this enzyme to undergo protein polymerisation, which may be associated to the proteins additional role as a structural protein in the formation of the mitochondrial capsule in sperm maturation. Two cysteine residues in addition to the catalytic cysteine were found to be associated with protein polymerisation in GPX4, which are not conserved in other human GPX enzymes thus accounting for the absence of a similar phenomenon.

Multiple sequence alignments of human GPX family members have revealed that GPX4 contains two sequence deletion regions that encode loop regions facilitating oligomerisation in other GPX members such as human GPX1 (Figure 1.11c), and thus concurs with the experimental observation that GPX4 is monomeric (Brigelius-Flohe et al., 1994). The absence of the two loop regions is suggested to make the active site more accessible to complex hydroperoxides and accounts for the broad substrate specificity of the enzyme, which is also observed in other PHGPX-like members including those of plants and trypanosomes. Human GPX4 lacks most of the glutathione-binding residues near the active site (Brigelius-Flohe et al., 1994), such as arginine residues in human GPX1, and is suggested to be able to use other thiols as a reductant other than glutathione.

The crystal structures of human GPX4 reveals the overall presence of a classic thioredoxin fold, consisting of a central beta sheet (comprising four beta strands) flanked by four alpha helices, with small additional N-terminal secondary structural elements (two alpha helices; two beta strands/hairpin structures). The tertiary structure has revealed the formation of the catalytic site consisting of the catalytic selenocysteine/cysteine (SeCys/Cys 73), glutamine (Gln 108) and tryptophan (Trp 163). These catalytic residues are located on a flat impression on the surface of the protein that extends into a surface exposed region containing positively charged residues, and is lined by apolar residues that maybe associated with the positioning of hydrophobic lipid peroxides (Scheerer et al., 2007). The human PHGPX enzyme has been shown to exhibit a classic selenocysteine GPX-type catalytic mechanism when coupled to glutathione, where catalytic residues glutamine and tryptophan play a role in activation and stabilisation of the selenocysteine residue via hydrogen bonding.

1.11 Thesis aims

The aims of this thesis were to design suitable constructs for *T. cruzi* GPXI, *T. cruzi* GPXII, *T. brucei* PxI and *T. brucei* PxII in order to produce suitable protein for structural investigations; to carry out structural characterisation of TcGPXI using X-ray crystallography and/or NMR spectroscopy; and to determine the molecular basis of the catalytic mechanism of TcGPXI.

Chapter Two

MATERIALS AND METHODS

Materials and methods

The following chapter outlines various experimental procedures that were used through the course of this study. All reagents are purchased from Sigma-Aldrich, Novagen or Invitrogen unless stated otherwise.

2.1 Sequence analysis

Sequence analysis of wild type $TcGPXI_{14-177}$, TbPxI, TbPxII, TcGPXIIa and TcGPXIIb constructs were carried out by submission of the amino acid sequences to the Interpro database (http://www.ebi.ac.uk/Tools/InterProScan/). In addition to this, sequences were submitted to the Pfam database in order to deduce the domains present, and to the Prosite database to investigate sequence-specific motifs/patterns associated to functional protein families.

2.2 Bacterial strains

The bacterial strains used in this study were competent *Escherichia coli* DH5α, *Escherichia coli* Rosetta(DE3)pLysS, *Escherichia coli* BL21(DE3), *Escherichia coli* C41(DE3), *Escherichia coli* C43 (DE3) and *Escherichia coli* Top10 (Invitrogen). The various strains were used to carry out test expressions in order to produce optimal yields of suitable protein for structural studies. The strain *Escherichia coli* Rosetta(DE3)pLysS was used for the expression of large protein amounts suitable for structural studies.

2.3 Growth medium and cultures

All media was freshly prepared with deionised water and sterilized by autoclaving. The growth media used to grow all *E. coli* bacterial strains harbouring the pProEx vector containing DNA coding various trypanosomal constructs, was supplemented with ampicillin at 100 μg/ml. The *E. coli* bacterial strain

Rosetta(DE3)pLysS was additionally supplemented with chloramphenicol at 33 μg/ml, respectively. Cell cultures supplemented with the appropriate antibiotic were grown in a shaking incubator (190 revolutions per minute (rpm) in a Stuart Scientific SI50 orbital incubator) at a temperature of 37 °C.

2.3.1 Luria-Bertani (LB) growth media

In order to cultivate *E. coli* strains, LB broth was prepared (10 g/L tryptone, 5 g/L yeast extract, 5 g/L NaCl) and sterilized by autoclaving. LB agar plates were produced by the addition of 15 g/L of agar to the LB media, which was poured evenly over petri dishes following sterilisation. Following site directed mutagenesis, the transformation of mutant clones were carried out using NZY^+ broth consisting of 10 g/L NZ amine (casein hydrolysate), 5 g/L yeast extract, 5 g/L NaCl. The pH was adjusted to 7.5 and the media sterilised, after which sterile filtered solutions of 12.5 ml/L of 1M $MgCl_2$, 12.5 ml/L of 1M $MgSO_4$ and 10ml/L of 2M glucose were added.

2.3.2 Minimal media

The expression of isotopically enriched proteins was carried out in minimal media. Minimal media was made with 6 g/L Na_2HPO_4, 3 g/L KH_2PO_4 and 0.5 g/L NaCL. For the expression of uniformly ^{15}N-labelled proteins, 1 g/L of ^{15}N-labelled $(NH_4)_2SO_4$ (Cambridge Isotope Laboratories) was added; for ^{13}C-labelled proteins, uniformly labelled glucose (Cambridge Isotope Laboratories) at 2 g/L was used. The pH of the minimal media solution was adjusted to pH 7.4 and the media was autoclaved. The following sterile filtered solutions were added per litre of minimal media prior to inoculation:

2ml 1M $MgCl_2$

10μl 1M $CaCl_2$

1ml 0.01M $FeSO_4$

1ml vitamin solution[†]

1ml micronutrient solution [††]

1ml ampicillin (100mg/ml stock solution)

1ml chloramphenicol (33 mg/ml stock solution)

† Vitamin Solution (1000x per litre)

0.4g choline chloride

0.5g folic acid

0.5g nicotinamide

1.0g myo-inositol

0.5g pyridoxal HCl

0.5g thiamine HCl

0.05g riboflavin

1.0g biotin

†† Micronutrient Solution (1000x per litre)

3μM ammonium molybate

400μM H_3BO_3

30 μM $CoCl_2$

10 μM $CuSO_4$

80 μM $MnCl_2$

10 μM $ZnSO_4$

2.4 Molecular cloning

2.4.1 Plasmid purification and preparation

In order to prepare plasmid DNA, fresh 10 ml *E. coli* DH5α bacterial cell cultures harbouring the required pProEx plasmid construct and ampicillin antibiotic were grown overnight at 37 °C shaking at 190 rpm (Stuart Scientific SI50 orbital incubator). The cells were pelleted by centrifugation and the plasmid purified using a QIAprep Spin Miniprep kit (Qiagen) using the manufacturer's protocol provided. The miniprep protocol uses alkaline lysis of the bacterial cells, and the lysate is

neutralized and adjusted for high-salt DNA binding conditions. The DNA is adsorbed onto a specialised silica membrane of the QIAprep column in the presence of high salt concentration, while RNA, cellular proteins and metabolites are not retained but found in flow-through. Wash steps facilitate the removal of endonucleases and salts, allowing the elution of high-quality plasmid DNA using the low-salt concentration EB buffer provided (EB contains 10mM Tris.Cl, pH 8.5).

2.4.2 Agarose gel electrophoresis

DNA fragments were separated and identified by agarose gel electrophoresis (1% or 1.5% agarose gels used). Agarose gels were made by dissolving agarose in TAE buffer (40mM 2-amino-2-hydroxymethyl-1,3-propanediol (Tris)-Acetate and 1 mM ethylene-diamine-tetraacetic acid (EDTA)) by heating. After the solution was cooled, ethidium bromide was added at a final concentration of 0.5 μg/ml in order to visualise the DNA. Samples of DNA were prepared by addition of DNA loading buffer (0.1% bromophenol blue, 20mM Tris and 20% glycerol). Molecular weight markers were used as standards (100bp or 1Kb ladders obtained from New England Biolabs (NEB). DNA agarose gel electrophoresis was performed at 100 V in the presence of TAE buffer and the resolved DNA fragments were visualised on a UV-transilluminator.

2.4.3 DNA extraction and purification from agarose gels

Resolved DNA fragments from agarose gel electrophoresis were extracted from the agarose gel using a clean scalpel, whereby bands of the appropriate size were excised. The QIAquick gel extraction kit (Qiagen) was used to purify the DNA from the gel using the manufacturer's protocol supplied. The QG buffer supplied with kit is used to solubilise the agarose gel slice and provide optimal conditions of pH and salt concentration in order to bind the DNA to the silica membrane. Unwanted impurities, such as salts, enzymes, free nucleotides, agarose, dyes, ethidium bromide, oils and detergents do not bind to the silica membrane and are

thus found in the flow through fraction. Wash steps facilitate the removal of salts and the purified DNA is eluted at a low salt concentration using the EB buffer.

2.4.4 Constructs and expression vectors

Oligonucelotide primers were designed in order to amplify gene fragments by polymerase chain reaction (PCR). Primers were designed for constructs varying the N-terminus of TcGPXI, TcGPXII, TbPxI and TbPxII (Chapter 3: Table 3.1). Primers were engineered with similar melting points, where forward primers (5'→ 3') were designed to incorporate the NarI (GGCGCC) restriction enzyme site and reverse (3'→ 5') primers designed to incorporate the HindIII restriction enzyme site (AAGCTT).

The amplified PCR products were resolved by agarose gel electrophoresis (1.5 % agarose gel), and the DNA purified by gel extraction (Section 2.4.3). The purified PCR products were sequentially digested with HindIII and NarI restriction enzymes at 37 °C for one hour and purified using the QIAquick PCR purification kit manufacturer's protocol to remove unwanted impurities and enzymes (Qiagen).

The pProEx expression vector containing TcGPXI sequence (residues 9-177) between restriction enzyme sites BamHI and HindIII (gift from Bloomsbury Centre for Structural Biology, BCSB), was used to obtain empty pProEx vector into which amplified PCR products were cloned with an N-terminal polyhistidine affinity tag and a tobacco etch virus (TEV) protease cleavage site to facilitate the removal of the affinity tag. The pProEx vector (4779 bp) contains various features including DNA coding for the β-lactamase gene conferring ampicillin resistance, a six histidine affinity tag, a seven amino acid spacer arm followed by a TEV protease recognition site and extensive multiple cloning site into which the gene of interest may be cloned. The vector also contains a trc promoter (hybrid promoter of the lac and trp promoters) and a $lacI^q$ gene for inducible expression of the cloned gene, resulting in a fusion protein containing a six histidine affinity tag at the amino terminus (facilitating protein purification with Ni-NTA resin) and TEV protease recognition site. Following expression and purification of the fusion protein, the affinity tag may be subsequently removed by the enzyme TEV protease.

Following plasmid purification of the starting plasmid construct from overnight 10ml *E. coli* DH5α bacterial cell cultures, the plasmid was sequentially digested with HindIII and NarI restriction enzymes to remove the $TcGPXI_{9-177}$ insert yielding empty pProEx vector. The digested plasmid vector was dephosphorylated by incubation for one hour at 37 °C with calf intestinal phosphatase (CIP). The digested vector was resolved using agarose gel electrophoresis (1% gel electrophoresis) and the empty vector DNA identified, extracted and purified from the agarose gel using the gel extraction kit (Qiagen).

2.4.5 Amplification of DNA using polymerase chain reaction (PCR)

PCR reactions were set up in 50 µl volumes consisting of: 1x *Pfu* DNA polymerase buffer (Promega) containing 200 mM Tris-HCl, 100mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM Mg_2SO4, 1mg/ml nuclease-free BSA, 1% Triton X-100) 200 µM dNTP mix (10mM each of dATP, dCTP, dGTP and dTTP in water, 1 µM upstream primer, 1 µM downstream primer, 0.5 µg template DNA and 1.25 U *Pfu* DNA polymerase enzyme. Various templates were used: for TcGPXI sequences a construct $TcGPXI_{9-177}$ cloned into the pProEx vector (gift from Bloomsbury Centre of Structural Biology) was used; for TcGPXII a plasmid construct of TcGPXII cloned into a pTrcHis vector (Invitrogen) (a kind gift from Dr. Shane Wilkinson, London School of Hygiene and Tropical Medicine (LSHTM)) was used; for the amplification of *T. brucei* sequences genomic DNA was used (a kind gift from Dr. Shane Wilkinson, LSHTM). PCR reactions were carried out using the Primus thermocycler (MWG) set to hold at 95 °C for 3 minutes for initial denaturation, then performing 30 cycles of 95 °C for 1 minute, annealing temperature 55 °C for 1 minute and a 72 °C extension for 1 minute/Kb to amplify the required DNA. Amplified PCR products were resolved on a 1.5 % agarose gel and purified following gel extraction.

2.4.6 Restriction enzyme digest

All restriction enzymes were obtained from New England Biolabs (NEB). PCR products were digested sequentially with NarI and HindIII restriction enzymes

at 37 °C for one hour using the manufacturer's protocol. The expression vector pProEx into which DNA fragments were cloned, were obtained by sequential digestion of the TcGPXI template construct ($TcGPXI_{9-177}$ cloned into pProEx vector) with NarI and HindIII, in order to remove the insert and generate empty pProEx vector. The plasmid vector was dephosphorylated using calf intestinal phosphatase (reaction carried out at 37 °C for one hour) to prevent re-ligation of the vector. Following digestion with restriction enzymes, DNA was purified by gel extraction using the Qiagen gel extraction kit protocol.

2.4.7 DNA ligation reactions

Ligation reactions were set up in 30 µl volumes containing 50ng of empty pProEx vector, 50 ng of insert, 1x T4 DNA ligation buffer (NEB) (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 10 mM dithioreitol (DTT), 1mM ATP, 25 µg/ml bovine serum albumin (BSA)) and 1µl of T4 DNA ligase (NEB). The ligation reaction mixture was left at 4 °C overnight. Ligations were tested for successful ligation of inserts by removing small samples and carrying out diagnostic restriction enzyme digests on reactions, followed by agarose gel electrophoresis. For TcGPXI constructs this involved digestion of a sample of the ligation with BamHI restriction enzyme, where successful clones no longer contained the BamHI site and were therefore uncleaved. For TcGPXII constructs, small samples of the ligations were digested with restriction enzymes XhoI and HindIII where those that were successful resulted in removal of an insert. Samples of the *T. brucei* ligations were digested only with HindIII restriction enzyme and compared to the control experiment of template $TcGPXI_{9-177}$ plasmid construct digested with HindIII, thus both should be linearised.

2.5 Transformation of bacterial cells

Aliquots of 50 µl *E.* coli DH5α cells (Invitrogen) were defrosted on ice for 30 minutes, after which 1 – 10 ng of ligation mixture or plasmid DNA was added. The mixture was then heat shocked at 42 °C for 20 seconds, returned to ice for a further two minutes and 950 µl of pre-warmed LB media was added. The transformations

were incubated at 37 °C for 1 hour shaking at 190 rpm (Stuart Scientific SI50 orbital incubator). Approximately 200 µl of the transformation was then plated on LB agar plates containing ampicillin antibiotic and the plates were incubated at 37 °C overnight. Transformation into *E. coli* Rosetta(DE3)pLysS (Novagen) was carried out using 20µl aliquots of cells defrosted on ice where following the addition of plasmid DNA, the transformation mixture was heat shocked at 42 °C for 45 seconds before returning it to ice for a further two minutes. Pre-warmed LB media (37 °C) was added (500 µl) and the mixture was incubated at 37 °C at 190 rpm (Stuart Scientific SI50 orbital incubator) for 1 hour, after which 80 µl was plated onto LB agar plates containing ampicillin and chloramphenicol antibiotics. The plates were left to incubate at 37 °C overnight. Transformation of BL21(DE3) cells was carried out as above with the exception that only ampicillin (100 µg/ml) was added to the LB agar plates.

2.6 Site directed mutagenesis of wild type TcGPXI

Three mutants of the cysteine residues (Cys48, Cys77 and Cys96) in the amino acid sequence of the wild type TcGPXI ($TcGPXI_{14\text{-}177}$) construct were created using the QuikChange Site-directed Mutagenesis kit from Stratagene.

Forward and reverse oligonucelotide primers were designed targeting the individual regions where specific point mutations were introduced to mutate each cysteine residue into a glycine residue (primer designs shown in Figure 3.18a). The mutagenesis PCR reaction was set up using the Quikchange Site Directed Mutagenesis Kite (Stratagene) and manufacturer's protocol, in a 50 µl volume containing 5 µl of 10x reaction buffer (100mM KCl, 100mM $(NH_4)_2SO_4$, 200mM Tris-HCl (pH 8.8), 20 mM $MgSO_4$, 1% Triton X-100, 1mg/ml nuclease-free bovine serum albumin), 125 ng forward primer, 125 ng reverse primer, 1 µl dNTP mix, 40 ng double stranded template (wild type $TcGPXI_{14\text{-}177}$ pProEx plasmid construct), 2.5 U *Pfu* Turbo DNA polymerase. The PCR reaction was carried out using the Primus thermocycler (MWG) set to 95 °C for 1minute, followed by 20 cycles of 95 °C for 30 seconds, 55 °C for 1 minute and 68 °C for 16 minutes. A 10 µl sample of the PCR reaction was then electrophoresed on a 1% DNA agarose gel to check the presence of the PCR product. The enzyme *Dpn* I was added (10 U), directly to the amplification

reaction and left to incubate at 37 °C for one hour in order to digest the non-mutated parental template DNA. Transformation into *E. coli* DH5α cells was then carried out, where 10 µl of the digested amplification mixture was added into 50 µl of the bacterial cells and left on ice for 30 minutes. The transformation mixture was heat shocked at 37 °C for 20 seconds before returning to ice for a further two minutes. 500 µl of pre-warmed NZY^+ media was added and the reaction mixture incubated at 37 °C at 190 rpm (Stuart Scientific SI50 orbital incubator) for 1 hour, after which 250 µl was plated onto ampicillin-containing LB agar plates. The plates were left to incubate at 37 °C overnight. Plasmid DNA of C48G, C77G and C96G TcGPXI mutants was purified from single colonies and sent for sequencing (MWG) for verification of mutations.

2.7 Glycerol stocks

A single colony from a LB agar plate containing *E. coli* bacterial cells harbouring the required plasmid, was inoculated into 10 ml of fresh LB media containing the appropriate antibiotic and grown overnight at 37 °C shaking at 190 rpm (Stuart Scientific SI50 orbital incubator). From the overnight cultures 1ml of cells was removed and placed into sterile cryotubes, into which 500 µl of 70% sterile glycerol was added. The glycerol stocks were stored at – 80 °C.

2.8 Small scale test expression

Small scale test expressions were carried out in 10 ml (for TcGPXI constructs) or 100 ml (for TcGPXII, TbPxI and TbPxII constructs) of LB media. A single recombinant colony was picked from LB agar plates (from fresh transformation of the appropriate *E. coli* strain) and used to inoculate 10ml overnight LB cultures containing appropriate antibiotic (ampicillin only, or combined with chloramphenicol). From the turbid overnight cell cultures 1ml or 0.5 ml was transferred into the 100 ml or 10 ml LB cultures, containing appropriate antibiotic (ampicillin at a final concentration of 100 µg/ml or both ampicillin and chloramphenicol (final concentration of chloramphenicol at 33 µg/ml) for *E. coli*

Rosetta(DE3)pLysS) and grown to an optical density (OD_{600}) of 0.6-0.8 at 37 °C shaking at 190 rpm (Stuart Scientific SI50 orbital incubator). The protein expression was induced by addition of isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 1mM. Protein expression was carried out at 30 °C for 3 hours, after which the cells were harvested by centrifugation at 6084 rcf for 15 minutes and then frozen at –20 °C until ready for purification.

2.9 Large scale protein expression

Large scale protein expression was routinely carried out in 2 or 3 litres of freshly prepared LB medium. A sample from the appropriate glycerol stock was streaked onto a fresh LB agar plate and incubated at 37 °C overnight. A single colony was picked from the LB agar plate and used to inoculate freshly prepared 10ml LB cultures inoculated with the appropriate concentrations of ampicillin and chloramphenicol antibiotics (for expression in *E. coli* Rosetta(DE3)pLysS). Overnight cultures were grown at 37 °C shaking at 190 rpm (Stuart Scientific SI50 orbital incubator). 10ml from the overnight cell cultures was then inoculated into freshly prepared 1 litre cultures of LB media containing the appropriate antibiotics. The cultures were grown at 37 °C, shaking at 190 rpm (Stuart Scientific SI50 orbital incubator) until an optical density (OD_{600}) of 0.6-0.8 was reached and then induced with IPTG at a final concentration of 1 mM. Protein expression was carried out at 30 °C for a further 3-4 hours and the cells then harvested by centrifugation at 6084 rcf for 15 minutes. The pelleted cells were stored at –20 °C until further use.

2.10 Protein purification

All protein purification was carried out at 4 °C. Pelleted cells containing expressed protein were defrosted on ice and resuspended in Buffer A (25 mM Tris pH 8.5, 200 mM NaCl, 10 mM β-mercaptoethanol) containing complete EDTA-free cocktail inhibitor (Roche). The cells were then lysed by sonication (10 x 30 second bursts with 30 second intervals). The lysate was then centrifuged at 26892 rcf for 45 minutes and the supernatant/lysate (soluble fraction) was separated from the pellet.

The lysate was incubated with 0.5-2 ml of pre-equilibrated (with Buffer A) Ni^{2+} - nitrilo-triacetic acid chelating beads (Ni-NTA agarose resin (Novagen)) rolling for 1 hour at 4 °C. The beads were then washed four times with 40 ml volumes of wash buffer (Buffer A containing 20 mM imidazole) and the protein eluted in elution buffer (Buffer A containing 250 mM imidazole). The eluted protein was analysed by SDS-PAGE analysis. In order to remove the affinity tag, his-tagged TEV protease was added to the recombinant protein at a ratio of 1:10 milligrams of TEV protease: recombinant protein, and the protein mixture was dialysed over forty-eight hours at 4 °C. Various dialysis buffers were used depending on the type of experiment for which protein was to be utilised. For crystallisation studies protein was dialysed into buffer containing 10 mM Tris pH 7.4 (at 4 °C) , 200 mM NaCl (in some cases 2mM DTT was added for studies of reduced forms of protein); for NMR experiments, dialysis was carried out in buffer containing 50 mM sodium phosphate pH 6.5, 300 mM NaCl (unless otherwise stated). Prior to the establishment of the standard purification procedure used in this study, extensive investigation into the requirement of reducing agent in the purification buffers was carried out. It was thus concluded that the presence of reducing agent was required in order to prevent over-oxidisation of the proteins and minimisation the formation of any oligomeric structures thus facilitating efficient cleavage of the histidine tag by TEV protease. Following incubation of the recombinant proteins with TEV protease the dialysed solution was passed over Ni-NTA agarose resin again so that TEV protease, cleaved affinity tag and any uncleaved protein was retained by the resin and removed from the purified protein. Purified cleaved protein was analysed by SDS-PAGE analysis (Section 2.16).

2.11 Buffer conditions for structural experiments

For crystallisation studies of oxidised forms of wild type TcGPXI, TcGPXIIa, TcGPXIIb, TbPxI, TbPxII protein was dialysed into buffer containing 10 mM Tris pH 7.4, 200 mM NaCl where 2 mM DTT was additionally added for crystallisation studies of the reduced form of wild type TcGPXI.

NMR experiments were carried out on wild type and mutant (C48G, C77G, C96G) TcGPXI proteins that were extensively dialysed into buffer of 50 mM sodium

phosphate pH 6.5, 300 mM NaCl (unless otherwise stated), where 10mM DTT was added for NMR investigation of the reduced wild type TcGPXI.

Kinetic studies were carried out following extensive dialysis of freshly prepared wild type and mutant (C48G, C77G, C96G) TcGPXI proteins into buffer containing 100mM Hepes pH 8.0, 0.5 mM EDTA, 10 mM β-mercaptoethanol (tryparedoxin-dependent assays) or into buffer containing 25 mM Tris pH 7.4, 200 mM NaCl, 10 mM β-mercaptoethanol (glutathione-dependent assays).

2.12 Size exclusion chromatography (SEC)

SEC was carried out using the Superdex100 (GE Healthcare, UK) pre-packed size-exclusion chromatography column and AKTA-FPLC system (Amersham Biosciences) at 4 °C. The column was pre-equilibrated with crystallisation buffer (10 mM Tris pH 7.42, 200 mM NaCl) and a protein sample (concentration of approximately 3 – 4 mg/ml) was injected onto the column at a flow rate of 0.5 ml/minute. The eluted protein was identified at a UV absorbance at 280 nm and appropriate fractions were retained and the presence of the target protein was verified by SDS-PAGE analysis.

2.13 Determination of protein concentrations

Protein concentrations were estimated using the Bio-Rad Protein assay based on the Bradford assay (Bradford, 1976), where differential colour change of a dye occurs in response to various concentrations of protein. The absorbance maximum for an acidic solution of Coomassie® Brilliant Blue G-250 dye shifts from 465 nm to 595 nm, upon binding to a protein. The dye primarily binds to arginine and aromatic amino acids where colour change can be measured by a spectrophotometer at 595nm. The protein concentration in a sample solution can be determined based on a standard reference curve generated by various concentrations of bovine serum albumin (BSA) as a standard. Absorbance measurements for a range of BSA concentrations within the linear range of the assay were made. The data was fitted to

a linear plot (using Microsoft Excel) and the equation describing the graph was used for subsequent protein concentration determinations.

2.14 Absorbance spectra and protein concentration determination

In addition to the Bio-Rad protein assay, protein concentrations were also determined using the UV-absorbance spectra of the proteins (measured in a 1ml quartz cuvette). Molar extinction coefficients of the various protein products were calculated and shown in Table 2.1 using the program Expasy ProtParam (Gasteiger et al., 2003). Protein concentrations were determined by application of the Beer-Lambert Law:

$$A_{280} = \varepsilon_{280} . C . l$$

where A_{280} is the absorbance of the protein at 280nm; ε_{280} is the molar extinction coefficient at 280 nm ($M^{-1}cm^{-1}$); C is the protein concentration (M); l is the path length (cm).

Construct	Molar extinction coefficient ($M^{-1}cm^{-1}$)
TcGPXI $_{14-177}$	17545
TbPxI $_{1-166}$	17545
TbPxII $_{5-169}$	17545
TcGPXIIa $_{69-225}$	17210
TcGPXIIb $_{5-160}$	17210

Table 2.1: Molar extinction coefficients for trypanosomal GPX enzymes
Molar extinction coefficients of protein products obtained in this study.

2.15 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

In order to resolve proteins according to their molecular size SDS-PAGE gels were made consisting of a 5 % stacking layer and 15% resolving layer. The SDS gel electrophoresis was carried out using the MiniProtean II gel apparatus (Biorad). Protein samples were prepared by the addition of 5 µl SDS loading dye (0.5 M Tris-HCl pH 6.8, 0.5% bromophenol blue, 10% SDS, 50% glycerol) to 10 µl of protein sample. The mixtures were heated at 95 °C for 5 minutes before loading onto the gel. Throughout this study SDS-PAGE analysis was carried out by loading 15 µl samples onto gels. Unstained Precision Plus Protein standards (Biorad) were used as molecular weight markers. Electrophoresis was carried out at a constant voltage of 100V through the stacking layer and then increased to 150V through the resolving layer. The 15% resolving layer of SDS-PAGE gels were made by combining 5ml of 30% acrylamide mix, 1.25 ml of 3M Tris pH 8.8, 0.1 ml of 10% SDS, 0.1 ml of 10 % ammonium persulphate, 3.55 ml of water and 0.004 ml of TEMED. The stacking layer was created by combining 0.5ml of 30% acrylamide mix, 0.38 ml of 1M Tris pH 6.8, 0.03 ml of 10% SDS, 0.03 ml of 10 % ammonium persulphate, 2.1 ml of water and 0.003 ml of TEMED. Following gel electrophoresis, the gel was stained with Coomassie dye (0.25 % (w/v) brilliant blue R Coomassie, 7 % (v/v) acetic acid, 52 % (v/v) H_2O, 40% (v/v) methanol). The resolved protein bands were then visualised by washing the gel in de-staining solution (40% (w/v) methanol, 10% (v/v) acetic acid and 50% (v/v) H_2O).

2.16 Nuclear magnetic resonance spectroscopy (NMR)

The nuclear magnetic resonance (NMR) spectra for [^{1}H, ^{15}N] HSQC and triple resonance experiments [HNCO, HNCA, HN(CO)CA, CBCA(CO)NH, HNCACB, HN(CA)CO, HA(CA)NH and HA(CACO)NH] were recorded at proton frequency of 500 or 600 MHz on the Varian Unityplus spectrometer, Varian INOVA spectrometer or Bruker Advance III spectrometer at 25 °C (unless otherwise stated) with established pulse sequences. Parameters are listed in table (Table 2.2 and Table 2.3). All NMR experiments were set up by Dr. Richard Harris using protein samples of specific concentration where stated. Sequence specific resonance assignments,

carried out by myself, were found by correlation of intra- and inter-residue resonances of Cα, Cβ and CO chemical shifts in order to create connected sequence segments which were then assigned using the probabilistic method of Grzesiek and Bax based upon the characteristic Cα and Cβ shift distributions of each residue type (Grzesiek and Bax 1993).

NMR data was processed using NMRPipe (Delaglio et al., 1995) and Azara software packages. Images of 2D spectra were produced using the program Plot2 from the Azara package. The 1D spectra images were created in NMRDraw.

	^{1}H		^{13}C		^{15}N		
Experiment	**Points**	**Sweep Width**	**Points**	**Sweep Width**	**Points**	**Sweep Width**	**Transients**
HSQC†	512	5000 (102.4 ms)			128	1700 (75.3 ms)	4
HSQC†	1024	8000 (128 ms)			128	1230 (104.1 ms)	8
HNCA†	1024	8000 (128 ms)	64	3270 (195.7 ms)	28	1230 (22.8 ms)	8
HN(CO)CA†	1024	8000 (128 ms)	64	3270 (195.7 ms)	28	1230 (22.8 ms)	8
HNCACB†	1024	8000 (128 ms)	64	7300 (8.8 ms)	28	1230 (22.8 ms)	20
CBCA(CO)NH†	1024	8000(128 ms)	45	7300 (6.1 ms)	28	1230 (22.8 ms)	16
HNCO†	512	5000 (102.4 ms)	48	1800 (26.7 ms)	32	1400 (22.9 ms)	8
HN(CA)CO††	512	4300 (119 ms)	48	2000 (24 ms)	32	1200 (26.7 ms)	32
HA(CA)NH††	512	4300 (119 ms)	48	2000 (24 ms)	28	1200 (23.3 ms)	16
HA(CACO)NH†	512	5000 (102.4 ms)	48	2200 (24 ms)	32	1400 (22.9 ms)	8

Table 2.2: NMR parameters used for oxidised wild type TcGPXI

The parameters used to acquire NMR data for oxidised wild type *T. cruzi* GPXI ($TcGPXI_{14-177}$).

† Spectra acquired on a Varian INOVA spectrometer operating at 1H frequency of 600 MHz.

†† Spectra acquired on a Varian UnityPlus spectrometer operating at 1H frequency of 500 MHz.

	^{1}H		^{13}C		^{15}N		
Experiment	**Points**	**Sweep Width**	**Points**	**Sweep Width**	**Points**	**Sweep Width**	**Transients**
HSQC *	512	4500 (113.8 ms)			128	1350 (94.8 ms)	8
HNCO*	1024	8000 (128 ms)	32	1500	32	1350 (23.7 ms)	16

Table 2.3: NMR parameters used for reduced wild type TcGPXI

Parameters used to acquire data for reduced wild type *T. cruzi* GPXI ($TcGPXI_{14-177}$)

* Spectra acquired on a Bruker Advance III spectrometer operating at 1H frequency of 600 MHz.

2.17 Kinetic investigations

All components for enzyme assays are from Sigma-Aldrich unless otherwise specified. Freshly prepared wild type TcGPXI, C48G, C77G and C96G mutant TcGPXI proteins under reducing conditions were used for all enzyme assays. To measure glutathione-dependent activity, enzyme assays were set up as described in previously published protocol (Wilkinson et al., 2000). In this protocol reducing equivalents from reduced nicotinamide adenine dinucleotide phosphate (NADPH) are used to reduce the hydroperoxide substrate, via glutathione and glutathione-dependent peroxidase enzymes (GPX). Glutathione reductase was included in the assay to facilitate electron transfer from NADPH to glutathione. This enzyme is otherwise absent in trypanosomal species and thus mimics the action of trypanothione and trypanothione reductase.

The enzymatic activity was measured spectrophotometrically following the oxidation of NADPH at 340nm at 30 °C. The standard reaction mixture carried out in a 1ml volume consisted of 100 mM Tris/HCl pH 7.4, 5mM EDTA, 0.2 mM β-NADPH, 1mM NaN_3, 3mM reduced glutathione, 0.1% (v/v) Triton X-100, 1.4 U *Saccharomyces cerevisiae* glutathione reductase and 2.2 μM of TcGPXI (or mutant TcGPXI). The basal levels of absorbance were obtained by measurements of oxidation in the absence of substrate where the reaction was followed for two minutes after initiation by the addition of t-butyl hydroperoxide substrate. The enzymatic activity was measured in duplicate reactions using a range of t-butyl hydroperoxide concentrations (3.6-50 μM).

The tryparedoxin-dependent enzymatic assays were carried out at 30 °C in 1ml volumes containing 50 mM Hepes pH 8.0, 0.5 mM EDTA, 200 μM NADPH, 0.5 μM trypanothione reductase*, 20 μM trypanothione (kind gift from Dr. Shane Wilkinson, LSHTM), 2 μM wild type TcGPXI (or mutant TcGPXI) enzyme, 1 μM recombinant *T. cruzi* tryparedoxin I ** and the substrate t-butyl hydroperoxide. All components except NADPH, trypanothione reductase and the TcGPXI enzyme were added as a blank sample for the assay. The NADPH was added and the absorbance checked prior to the addition of trypanothione reductase. The reaction was initiated by addition of TcGPXI enzyme and followed spectrophotometrically over a period of two minutes. Basal levels of absorbance of NADPH oxidation in the absence of

TcGPXI enzyme were monitored and the enzymatic activity measured using various t-butyl hydroperoxide substrate concentrations (0-100 μM).

* *Crithidia fasciculata* trypanothione reductase (CfTR) was kindly produced by Dr. Shane Wilkinson (LSHTM) and Dr. Syeed Hussain (UCL) using established protocol. The SDS-PAGE analysis of trypanothione reductase fractions shows bands corresponding to trypanothione reductase approximately 52 kDa (Figure 2.1a).

** Recombinant Tryparedoxin I

Recombinant *T. cruzi* tryparedoxin I was produced by myself, using a construct of *T. cruzi* tryparedoxin I DNA cloned into a pTrcHis expression vector (kind gift from Dr. Shane Wilkinson, LSHTM). The plasmid was transformed into *E. coli* Top 10 cells and plated onto ampicillin-containing LB agar plates. A single colony was used to inoculate a 10 ml overnight LB culture containing ampicillin antibiotic. The 10 ml overnight cell cultures were inoculated into fresh, one litre LB media containing ampicillin (100 μg/ml). The cells were grown at 37 °C, shaking at 190 rpm (Stuart Scientific SI50 orbital incubator) until the optical density (OD_{600nm}) reached 0.6 - 0.8. Protein was expressed by induction with IPTG, at a final concentration of 1mM for 3-4 hours at 37 °C, after which the cells were harvested and stored at -20 °C until further use. Protein was purified in the same manner as detailed earlier (Section 2.11), using batch purification methods with Ni-NTA resin. The recombinant protein was extensively dialysed in buffer containing 100mM Hepes pH 8.0, 0.5 mM EDTA for enzymatic assays. Two litre cell cultures of *E. coli* Top10 harbouring the *T. cruzi* tryparedoxin I plasmid construct consistently yielded approximately 9-10 mg of protein (0.97 mg/ml). The purified protein was analysed by SDS-PAGE (Figure 2.1b) where bands corresponding to purified recombinant *T. cruzi* tryparedoxin I were present at approximately 18 kDa.

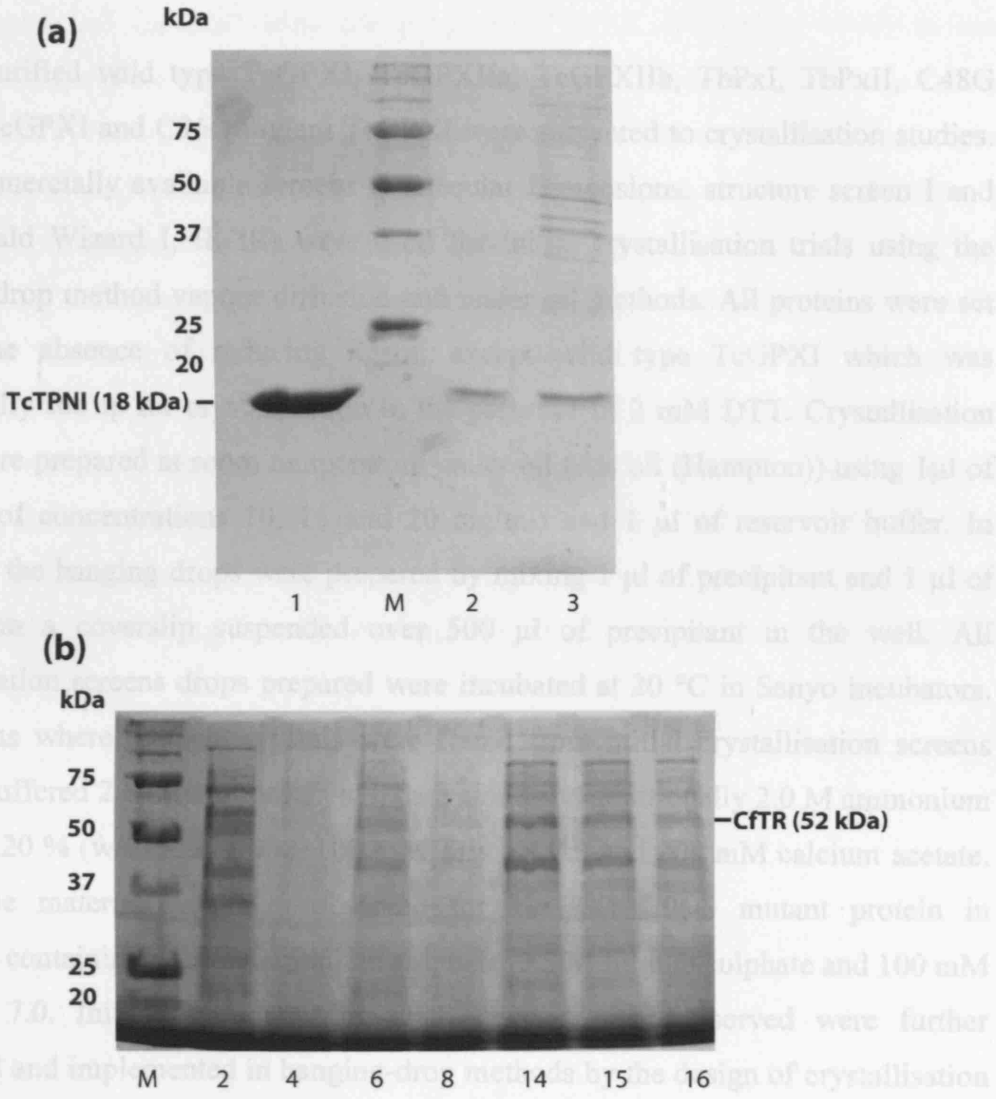


Figure 2.1: SDS PAGE analysis of *T. cruzi* tryparedoxin I and *C. fasciculata* trypanothione reductase

(a) SDS PAGE analysis of purification of recombinant *T. cruzi* tryparedoxin I (TcTPNI). Lane 1: purified *T. cruzi* tryparedoxin I (TcTPNI) protein at approximately 18 kDa; Lane 2: Ni-NTA resin from purification; Lane 3: first wash fraction.

(b) Fractions used for kinetic studies from purification of trypanothione reductase from *C. fasciculata*. The band corresponding to trypanothione reductase (CfTR) is identified at approximately 52 kDa.

2.18 Crystallisation studies

Purified wild type TcGPXI, TcGPXIIa, TcGPXIIb, TbPxI, TbPxII, C48G mutant TcGPXI and C96G mutant TcGPXI were subjected to crystallisation studies. The commercially available screens (Molecular Dimensions: structure screen I and II; Emerald Wizard I, II, III) were used for initial crystallisation trials using the hanging drop method vapour diffusion and under oil methods. All proteins were set up in the absence of reducing agent, except wild type TcGPXI which was additionally set up for crystallisation in the presence of 2 mM DTT. Crystallisation drops were prepared at room temperature under oil (Als oil (Hampton)) using 1µl of protein (of concentrations 10, 15 and 20 mg/ml) and 1 µl of reservoir buffer. In addition, the hanging drops were prepared by mixing 1 µl of precipitant and 1 µl of protein on a coverslip suspended over 500 µl of precipitant in the well. All crystallisation screens drops prepared were incubated at 20 °C in Sanyo incubators. Conditions where suitable crystals were found from initial crystallisation screens were unbuffered 2.0 M ammonium sulphate alone and additionally 2.0 M ammonium sulphate, 20 % (w/v) PEG-3000, 100 mM Tris pH 7.0 and 200 mM calcium acetate. Crystalline material was also observed for TcGPXI C96G mutant protein in condition containing 2.0 M ammonium sulphate, 0.2 M lithium sulphate and 100 mM Tris pH 7.0. Initial conditions in which crystals were observed were further optimised and implemented in hanging-drop methods by the design of crystallisation screens where ammonium sulphate concentrations and pH were varied. Additional screens were also set up where the concentration of unbuffered ammonium sulphate was varied. Crystals suitable for diffraction were found in hanging drop trays at protein concentrations of approximately 15 and 20 mg/ml in crystallisation conditions of unbuffered 1.8 M, 1.9 M and 2.0 M ammonium sulphate.

2.18.1 Data collection, processing and refinement

Wild type TcGPXI crystals that were found to be suitable for data collection were first submerged into a cryoprotectant of 20 % sterile glycerol before mounting onto the gonoimeter and positioning in the path of the X-ray beam. The data was obtained in-house using an R-AXISIV image plate detector system and Osmic

mirrors. 180 ° of data was collected over 1 ° oscillations and images were scaled, data merged and indexed using the program d*TREK (Pflugrath, 1999) in the CrystalClear software suite (all X-ray data was kindly collected and processed with the help of Dr. Gary Parkinson at the London School of Pharmacy). Iterative stages of refinement and manual fitting were carried out using maximum likelihood restrained refinement, with simple scaling in REFMAC (Murshudov et al., 1997) of the CCP4 crystallography suite and simulated annealing torsion angle dynamics with maximum likelihood function in CNS (Rice and Brunger, 1994), where the temperature of the system was heated to 2500K with slow cooling stages (refinement in CNS was kindly carried out with the help of Dr. Snezana Djordjevic). Sigma weighted electron density maps were calculated in REFMAC or COOT (Emsley and Cowtan, 2004). Between stages of refinement and map calculation, real space refinement/manual fitting and model building were carried out in the graphics program of COOT.

Chapter Three

Results

CLONING, PROTEIN PURIFICATION AND CRYSTALLISATION

3. 0 Design of gene constructs for the production of trypanosomal glutathione-dependent peroxidase enzymes

The full length sequences of TcGPXI, TcGPXII and TbPxIII consist of 177, 160 and 176 amino acid residues respectively. Biochemical and functional characterisation of these enzymes has been previously carried out (Hillebrand et al., 2003; Wilkinson et al., 2000a; Wilkinson et al., 2002b).

The initial aim of this study was to design and clone suitable TcGPXI constructs by established molecular cloning methods, whose transformation and protein expression in *E. coli* bacterial expression host strain would lead to the production of optimal protein products for detailed structural analysis. In addition to this, suitable constructs were also designed for *T. brucei* PxI, PxII and *T. cruzi* GPXII which were also used for structural studies.

3.1 Construct designs and sub-cloning

Folded proteins can contain flexible, structurally disordered regions at their N- or C-terminal ends, which may affect the production of soluble, stable protein and can hinder structural studies, as highly flexible terminal regions can affect the ability of a protein to crystallise (Kundrot, 2004).

At the time of this work the closest protein sequence homologue to trypanosomal GPX enzymes was selenocysteine-containing bovine cellular GPX1 (bovine cGPXI) exhibiting 33% sequence identity, the crystal structure of which has been solved (PDB code 1GP1; (Epp et al., 1983). The sequence of this protein was used to carry out a structure-based sequence alignment and design of DNA constructs that would provide the desired protein products. Subsequently the crystal structures of other TcGPXI homologues human PHGPX4 (sequence identity to TcGPXI 42%; PDB codes 2GS3 and 2OBI), and *Populus trichocarpa x deltoides* GPX5 (sequence identity to TcGPXI 48%; PDB code 2P5R/2P5Q) have been determined.

A multiple sequence alignment (CLUSTALW) was generated (Figure 3.1) in order to design N-terminus primers for TcGPXI, TcGPXII, TbPxI and TbPxII correlating to the N-terminal starting residue of the amino acid sequence of the

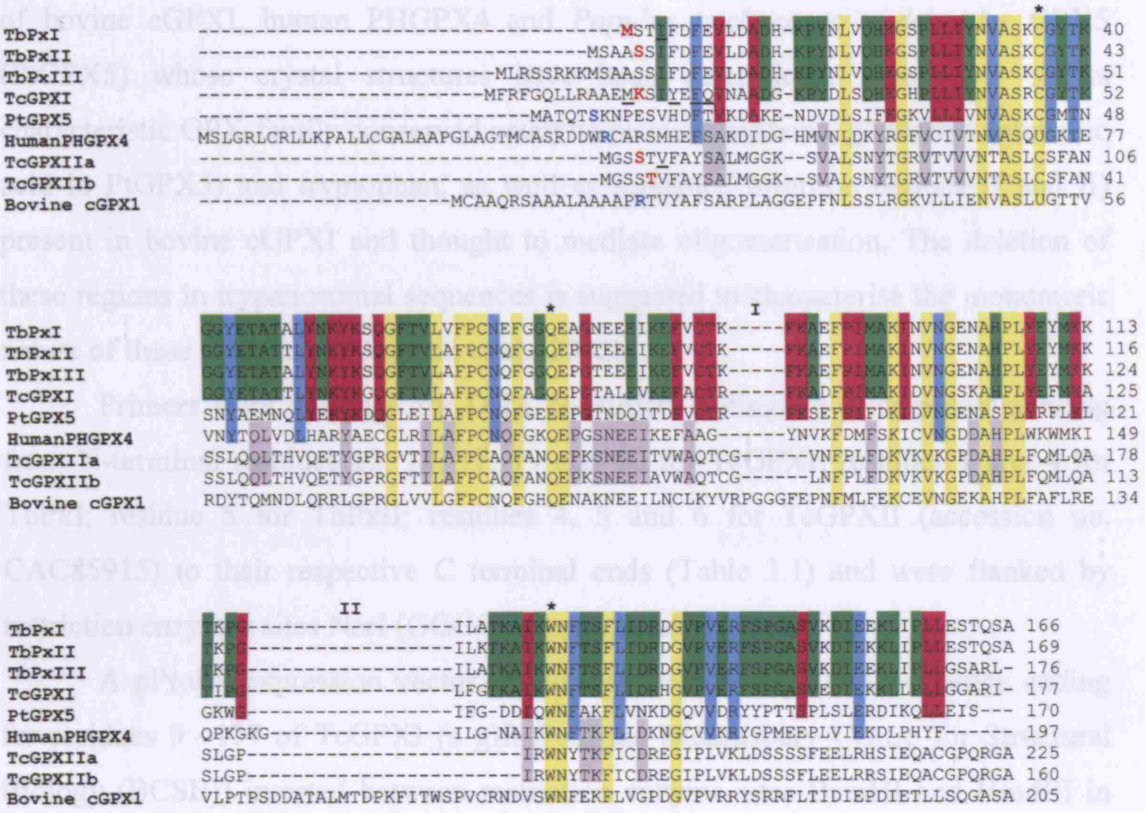


Figure 3.1: Structure-based sequence cloning

Multiple sequence alignment of trypanosomal GPX sequences and homologues. Sequence conservation is highlighted: conservation between all sequences (yellow); between TbPxI, TbPxII, TcGPXI (green); between TbPxI, TbPxII, TcGPXI, PtGPX5 (magenta); between TbPxI, TbPxII, TcGPXI, PtGPX5, human PHGPX4 (blue); between TcGPXIIa, TcGPXIIb, human PHGPX4 (grey). Residues contributing to the conserved GPX family catalytic triad (cysteine/selenocysteine, glutamine and tryptophan) are denoted *. Insertion regions I /II present in bovine cGPXI are indicated. N-terminal residues corresponding to construct designs (underlined) and of constructs carried forward (red) for structural studies ($TcGXPI_{14-177}$, $TbPxI_{1-166}$, $TbPxII_{5-169}$, TcGPXIIa $_{69-225}$, TcGPXIIb $_{5-160}$) are indicated. Sequences and accession no. (given in parentheses) used to generate the alignment: **TcGPXI**, *T. cruzi* GPXI (CAC85914); **TbPxI**, *T. brucei* PxI (CAC83347); **TbPxII**, *T. brucei* PxII (CAC83348); TcGPXIIa, *T. cruzi* GPXII (XP_807101); **TcGPXIIb**, *T. cruzi* GPXII (XP_814034). N-terminus residues (blue) of constructs corresponding to GPX sequences whose crystal structures have been solved (**PtGPX5**, *P. trichocarpa x deltoides* GPX5 (ABN59534; PDB 2P5R/2P5Q); Human PHGPX4, *H. sapiens* phospholipid GPX-4 (P36969; PDB 2GS3); Bovine cGPXI, *B. taurus* cellular GPX-1 (P00435, PDB 1GP1) are shown.

crystal structure 1GP1. The sequence alignment highlights various degrees of sequence conservation between trypanosomal sequences and homologous sequences of bovine cGPXI, human PHGPX4 and *Populus trichocarpa x deltoides* GPX5 (PtGPX5) whose crystal structures have been determined. Figure 3.1 shows characteristic GPX-family conserved catalytic residues cysteine, glutamine (glutamic acid in PtGPX5) and tryptophan, as well as sequence insertion regions (I and II) present in bovine cGPXI and thought to mediate oligomerisation. The deletion of these regions in trypanosomal sequences is suggested to characterise the monomeric nature of these proteins (Wilkinson et al., 2002b).

Primers were designed for DNA constructs coding for protein commencing from N-terminal residues 13, 14, 17, 19 and 20 for TcGPXI; residues 1 and 4 for TbPxI; residue 5 for TbPxII; residues 4, 5 and 6 for TcGPXII (accession no. CAC85915) to their respective C terminal ends (Table 3.1) and were flanked by restriction enzymes sites NarI (GGCGCC) and HindIII (AAGCTT).

A pProEx expression vector (Invitrogen) containing DNA sequence coding for residues 9 –177 of TcGPXI (a gift from the Bloomsbury Centre for Structural Biology (BCSB)) inserted between restriction enzyme sites BamHI and HindIII in the multiple cloning site, was used to obtain empty vector into which DNA inserts could be cloned. In order to obtain empty pProEx vector (*i.e.* remove DNA insert coding for $TcGPXI_{9\text{-}177}$), the plasmid was sequentially digested with NarI (GGCGCC) and HindIII (AAGCTT) restriction enzymes at 37 °C for one hour. The linearised vector was additionally dephosphorylated by incubation at 37 °C for one hour with calf intestinal phosphatase to prevent re-ligation of the digested vector. The digested vector was visualised on a 1% agarose gel stained with ethidium bromide and subsequently purified from the agarose gel (Section 2.4.2).

The DNA coding sequence for the TcGPXI protein constructs/variants were amplified, using the various primers shown in Table 3.1a, by Polymerase Chain Reaction (PCR) with the pProEx plasmid (Invitrogen), containing $TcGPXI_{9\text{-}177}$ sequence (from BCSB), as a template. The PCR reactions were carried out as described in Section 2.4.5 and subsequent PCR reaction products were visualised on a 1.5% agarose gel stained with ethidium bromide. The DNA of the PCR products was purified from the agarose gel (Section 2.4.3) and digested with restriction enzymes NarI and HindIII. The PCR products (inserts) were then ligated into the digested empty pProEx vector (Section 2.4.7) and the resultant clones were

(a)

CONSTRUCT	N-term starting residue	PRIMER
TcGPXI	13	TAGTAGTGGCGCCATGAAAAGCATCTAC
TcGPXI	14	TAGTAGTGGCGCCAAAAGCATCTAC
TcGPXI	17	TAGTAGTGGCGCCTACGAATTTCAGGTG
TcGPXI	19	TAGTAGTGGCGCCTTTCAGGTGAACGC
TcGPXI	20	TAGTAGTGGCGCCCAGGTGAAC
(Reverse)	-	CAAAACAGCCAAGCTTTCAAATCCTAGC

(b)

CONSTRUCT	N-term starting residue	PRIMER
TbPxI	1	TAGTAGTGGCGCCATGTCTACAATCTTTGACTTTGAG
TbPxI	4	TAGTAGTGGCGCCATCTTTGACTTTGAGGTGCTTGACGCCG
TbPxII	5	TAGTAGTGGCGCC TCT TCA ATC TTT GAC TTT GAG GTG
(Reverse)	-	CAAAACAGCCAAGCTTTCACGCGCTCTGCGTGC

(c)

CONSTRUCT	N-term starting residue	PRIMER
TcGPXIIa	4	TAGTAGTGGCGCCAGCACGGTGTTCGCGTACTCCGCGC
TcGPXIIb	5	TAGTAGTGGCGCCACGGTGTTCGCGTACTCCGCGCTGATG
TcGPXIIc	6	TAGTAGTGGCGCCGTGTTCGCGTACTCCGCGCTGATGGG
(Reverse)	-	CAAAACAGCCAAGCTTTCATGCACCCCGTTGC

Table 3.1: Primer designs for structure-based DNA constructs design

The 5'→ 3' (forward) and 3'→5' (reverse) primer designs for constructs varying at the N-terminus for TcGPXI (a), TbPxI/TbPxII (b) and TcGPXII (c). Restriction enzyme sites for NarI (grey) and HindIII (yellow) are highlighted.

transformed into competent *E. coli* DH5α cells. Plasmids were purified from several colonies of each construct and subjected to a diagnostic test whereby plasmids were digested with the restriction enzyme BamHI and compared to a control of the original TcGPXI template linearised with BamHI. Those constructs containing the newly designed TcGPXI inserts no longer contained the BamHI restriction enzyme site present and therefore could not be linearised by this enzyme. Subsequently the resultant clones were sent for sequencing to the company MWG.

The resultant proteins contained an N-terminal hexahistidine affinity tag to facilitate protein purification, and a *Tobacco Etch Virus* (TEV) protease cleavage site for the removal of the affinity tag. Following cleavage with TEV protease only an additional plasmid-derived glycine and alanine residues would remain at the N-terminus of each TcGPXI variant construct.

The TcGPXII, TbPxI and TbPxII constructs, using primer designs shown in Table 3.1b-c, were amplified by PCR in the same manner as TcGPXI but using different templates. The plasmid pTrcHis (Invitrogen) containing DNA coding for full length TcGPXII (amplified from CL Brener genomic DNA) was used as a template to amplify TcGPXII variants by PCR, and *T. brucei* genomic DNA was used as template in the PCR reactions to amplify TbPxI and TbPxII variants (both templates were gifts from Dr. Shane Wilkinson, LSHTM). The resultant PCR products of TcGPXII, TbPxI and TbPxII variants were sub-cloned into pProEx plasmid digested with NarI and HindIII restriction enzymes (empty vector) in a protocol analogous to that carried out for TcGPXI. In order to assess whether successful ligation of the respective inserts into the empty pProEx vector had been achieved diagnostic tests using restriction enzymes XhoI and HindIII were performed. The DNA sequence coding for TcGPXII variant constructs contain both Xho1 and HindIII restriction enzyme sites, whereas the *T. brucei* constructs contain only the HindIII restriction site. Therefore, successfully cloned constructs of TcGPXII digested with XhoI and HindIII resulted in removal of an insert out of the vector when visualised on an ethidium bromide stained, 1% agarose gel. The *T. brucei* constructs that were successfully cloned resulted in a linearised plasmid upon digestion with only HindIII, similar to the control experiment (the template $TcGPXI_{9\text{-}177}$ construct was also linearised upon digestion with HindIII) (Figure 3.2).

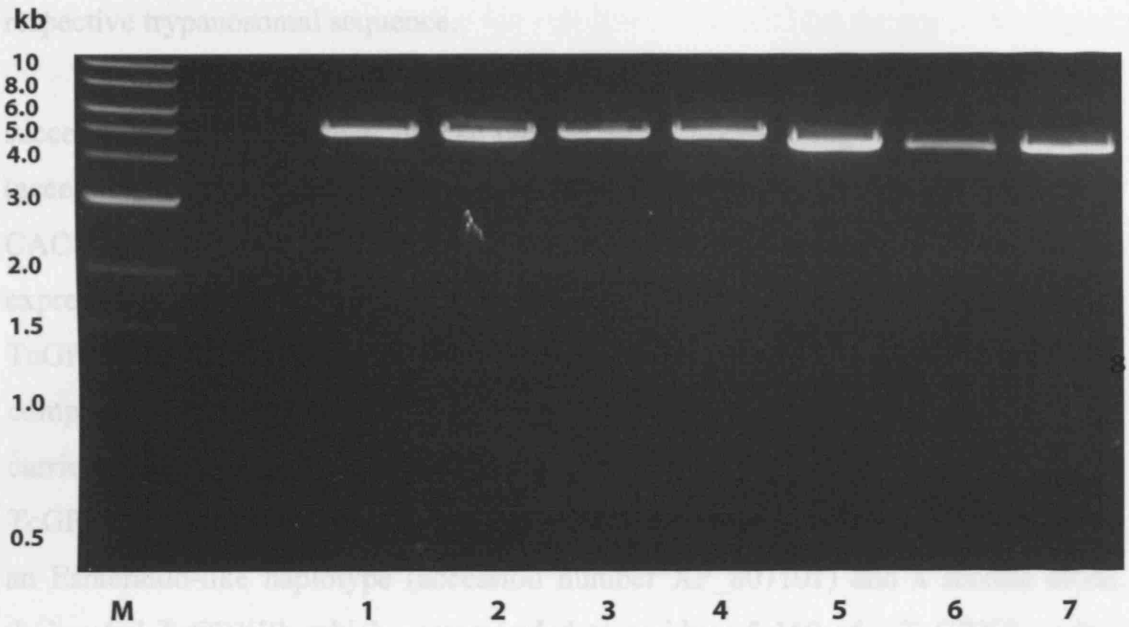


Figure 3.2: Diagnostic digest for cloning of *T. brucei* and *T. cruzi* GPXII constructs

DNA agarose gel (1%) stained with ethidium bromide. Lanes 1- 4 show cloned constructs in pProEx vectors linearised upon digestion with HindIII restriction enzyme. Lanes 5-7 shows cloned constructs digested with restriction enzymes Xho1 and HindIII resulting in partial removal of the cloned inserts.

Lanes: (1) template pProEx vector containing TcGPXI$_{9\text{-}177}$ (control); (2) TbPxI$_{1\text{-}166}$; (3) TbPxI$_{4\text{-}166}$; (4) TbPxII$_{5\text{-}169}$; (5) TcGPXIIa $_{69\text{-}225}$; (6) TcGPXIIa $_{71\text{-}225}$; (7) TcGPXIIb $_{5\text{-}160}$. Molecular weight marker is denoted (M).

The resultant clones were confirmed by sequencing via the company MWG. The resultant proteins, similar to TcGPXI, contain an N-terminal hexahistidine affinity tag followed by a TEV protease cleavage tag. Upon removal of the affinity tag by TEV protease only a glycine and alanine would remain prior to the start of the respective trypanosomal sequence.

The results of sequencing showed all constructs for TcGPXI had been successfully cloned. From *T. brucei*, two variants corresponding to full length TbPxI (accession number CAC83347) and residues 5-169 of TbPxII (accession number CAC83348) were cloned successfully and chosen to be carried forward for protein expression and purification. The sequencing results of the clones generated for TcGPXII sequences revealed that the template plasmid used in the PCR reaction comprised a mixed population of plasmids. Two resulting clones were chosen to be carried forward for the expression and purification of protein: clone designated TcGPXIIa which corresponded to residues 69-225 of a TcGPXII variant shown to be an Esmeraldo-like haplotype (accession number XP_807101) and a second clone, designated TcGPXIIb which corresponded to residues 5-160 of a TcGPXII variant shown to be a non-Esmeraldo-like haplotype (accession number XP_814034). The observed variations in the amino acid sequences of the TcGPXII clones are coherent with the fact that the CL Brener strain is a hybrid strain. The *T. cruzi* taxon consists of groups I and II according to their infectivity and host range. The taxon also contains additional groups yet to be designated. *T. cruzi* II is further subdivided into related groups IIa, IIb, IIc, IId and IIe. The strain CL Brener, is the reference strain for the genome project and although it is a member of subgroup IIe, it is also a hybrid of subgroup IIb and IIc (which is also a hybrid derived from *T. cruzi* I). The CL Brener strain is found to contain sequences representing two different haplotypes that may be differentiated by one of the haplotypes showing greater similarity to the Esmeraldo genome (member of the progenitor subgroup IIb) (El-Sayed et al., 2005a).

3.2 Small scale test expression of TcGPXI constructs

The trypanosomal genes were cloned into a pProEx expression plasmid (Invitrogen) containing the ampicillin resistance gene. The pProEx vector contains a trc promotor and $lacI^q$ (encodes high levels of lac repressor) gene that enables

inducible expression of a cloned gene with the highly stable synthetic analogue of lactose isopropyl-beta-D-thiogalactopyranoside (IPTG) once transformed into an *Escherichia coli* host expression strain. The trc promoter is a hybrid promoter of the lac and trp promoters and is a stronger promoter in comparison to the lac promoter. Upon induction, IPTG prevents the lac repressor from binding the promoter region, thus allowing transcription of the cloned gene.

The trypanosomal clones were initially transformed into the *E. coli* DH5α bacterial strain routinely used for subcloning into plasmid vectors for high-quality plasmid preparation. To initially assess the protein expression of the five TcGPXI constructs ($TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$, $TcGPXI_{17\text{-}177}$, $TcGPXI_{18\text{-}177}$ and $TcGPXI_{19\text{-}177}$) small scale test expressions were carried out in *E. coli* DH5α cell line, using 10ml Luria-Bertani (LB) cultures. The cultures were grown at 37 °C to an optical density of 0.6-0.8 (OD_{600nm}) and recombinant protein was expressed by the addition of IPTG to a final concentration of 1mM (Section 2.9). Protein expression was carried out at 30 °C for 3 hours after which cells were harvested by centrifugation at 26892 rcf. The cells were sonicated and purified as described in Section 2.11. The resultant recombinant proteins were visualised using SDS-PAGE analysis (20 μl of each protein sample was loaded onto the gel) where constructs $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$ and $TcGPXI_{18\text{-}177}$ were found to produce highest yields and selected to be carried forward for protein purification.

3.3 Selection of a wild type TcGPXI construct

The cloning of the chosen TcGPXI variants into the pProEx plasmid vector facilitates the expression of recombinant TcGPXI proteins containing a six consecutive histidine affinity tag, a 7-amino acid spacer arm and a TEV protease recognition site prior to the N-terminus of the respective TcGPXI protein. The histidine tag binds to Ni^{2+} cations, which are immobilized on a His-binding metal chelation resin (Ni-NTA resin). The metal is held by chelation with reactive groups covalently attached to a solid support. The Ni-NTA His-bind resin use nitriloacetic acid (NTA) as a chelator, which has four sites available for interaction with metal ions leaving two sites free to interact with histidine residues of the affinity tag of the recombinant protein. This allows the purification of the protein of interest. All

protein purification procedures in this study were performed using Immobilised Metal Ion Affinity Chromatography using a batch method carried out at 4 °C. Detailed purification procedures are given in materials methods but a brief overview is given in Figure 3.3.

Following selection of the three clones ($TcGPXI_{13-177}$, $TcGPXI_{14-177}$ and $TcGPXI_{18-177}$), large scale purification of these clones, including the original template vector ($TcGPXI_{9-177}$), was carried out to investigate and select which of the designed constructs gave a maximal yield of stable and soluble protein suitable for structural studies.

Large scale cell cultures (0.5 litre) were grown to an optimal density of 0.6-0.8 ($OD_{600-800nm}$) at 37 °C, followed by induction with IPTG at a final concentration of 1mM. The protein was expressed at 30 °C for 3-4 hours. Cells were then harvested ready for purification of the recombinant proteins. The harvested *E. coli* DH5α cells were lysed by sonication and the TcGPXI proteins were retained by Ni-NTA resin, and any non-specific binding of other proteins was eliminated by washing with 20mM imidazole buffer. The TcGPXI proteins were retrieved by elution with buffer containing 250mM imidazole. Under increasing imidazole concentration the His-tag TcGPXI fusion proteins dissociate from the Ni-NTA resin as they can no longer compete for binding sites, therefore the recombinant protein is eluted. TEV protease (containing N-terminal histidine tag) was added to the elutions in order to cleave the histidine tags from the TcGPXI proteins. To remove the cleaved histidine tag, uncleaved tag and TEV protease the solution was passed over Ni-NTA resin again, and unbound cleaved protein was collected. The addition of up to 10 mM β-mercaptoethanol was applied throughout purification buffers for the prevention of disulphide bond formation.

The purified TcGPXI proteins ($TcGPXI_{13-177}$, $TcGPXI_{14-177}$, $TcGPXI_{18-177}$ and $TcGPXI_{9-177}$) were analysed by SDS-PAGE. Figure 3.4 shows bands at approximately 20-22 kDa corresponding the eluted His-tagged TcGPXI proteins and samples of the Ni-NTA resin after elution of protein, where some protein still remains bound due to non-specific binding. From observation of the band intensities corresponding to the His-tagged eluted proteins, it may be seen that constructs $TcGPXI_{13-177}$ and $TcGPXI_{14-177}$ expressed greater protein amounts than $TcGPXI_{18-177}$ and the original construct $TcGPXI_{9-177}$. The proteins obtained from TcGPXI constructs are estimated to be > 90% pure based on SDS-PAGE analysis.

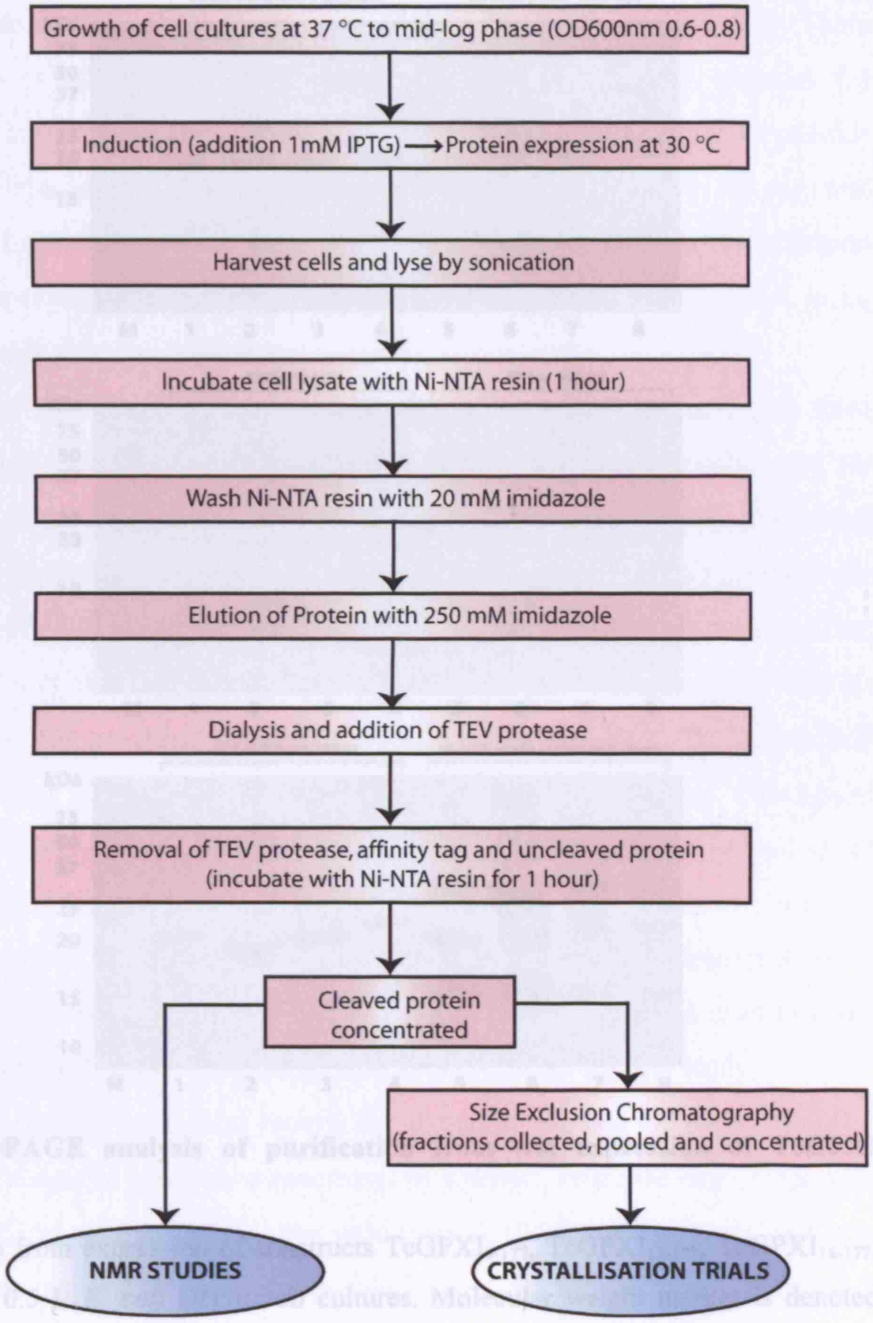


Figure 3.3: Flow diagram of purification procedure

Flow chart summarising sequential stages of *E. coli* cell growth, protein expression and protein purification, for structural studies. SDS-PAGE is used to analyse the resultant protein.

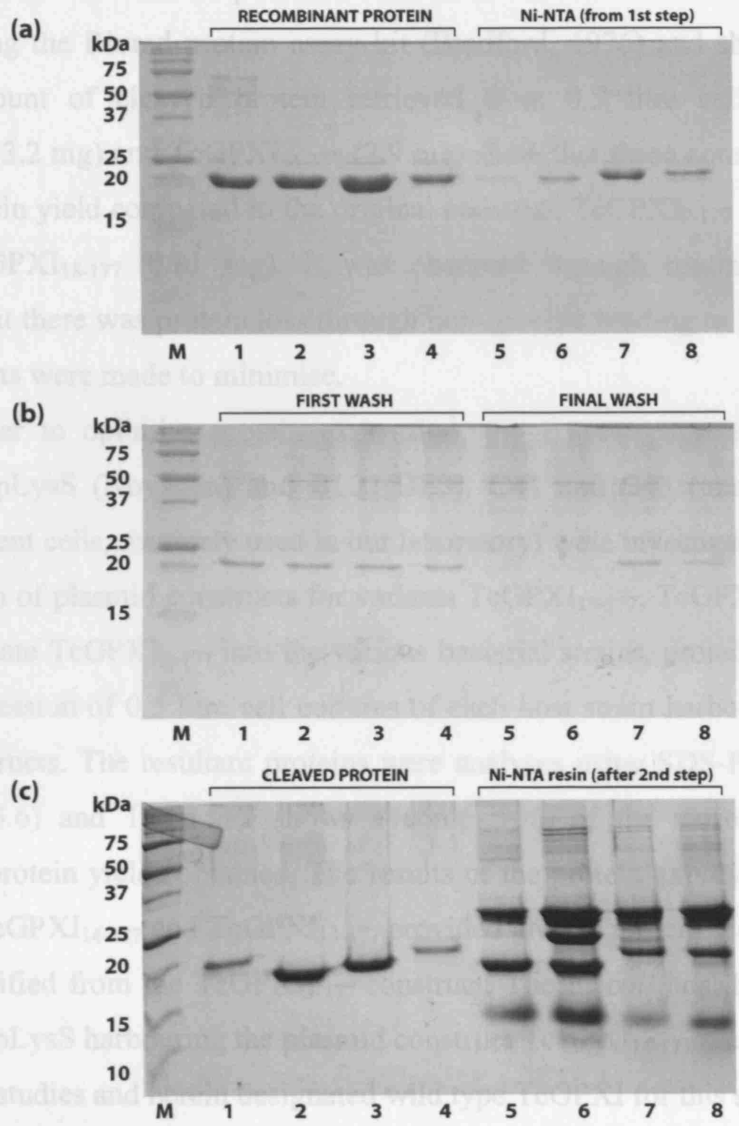


Figure 3.4: SDS-PAGE analysis of purification from test expression of TcGPXI constructs

Protein purification from expression of constructs $TcGPXI_{9-177}$, $TcGPXI_{13-177}$, $TcGPXI_{14-177}$, $TcGPXI_{18-177}$ from 0.5 L *E. coli* DH5α cell cultures. Molecular weight marker is denoted (M). Equal volumes of each sample (treated equally) were loaded onto the gel.

(a) Recombinant His-tagged protein of $TcGPXI_{9-177}$, $TcGPXI_{13-177}$, $TcGPXI_{14-177}$, $TcGPXI_{18-177}$ (lanes 1-4 respectively) and Ni-NTA resin (lanes 5-8 respectively).

(b) First wash fractions for $TcGPXI_{9-177}$, $TcGPXI_{13-177}$, $TcGPXI_{14-177}$, $TcGPXI_{18-177}$ (lanes 1- 4 respectively) and final wash fractions (lanes 5-8 respectively).

(c) Eluted cleaved protein not bound to Ni-NTA resin (affinity tag removed) for $TcGPXI_{9-177}$, $TcGPXI_{13-177}$, $TcGPXI_{14-177}$, $TcGPXI_{18-177}$ (lanes 1- 4 respectively) and Ni-NTA resin used to remove TEV protease, uncleaved protein and affinity tag (lanes 5-8).

The total amount of soluble protein obtained from 0.5 litre cell cultures was estimated using the Biorad protein assay kit (Bradford, 1976) and shown in Table 3.2. The amount of cleaved protein retrieved from 0.5 litre cell cultures for $TcGPXI_{13-177}$ (3.2 mg) and $TcGPXI_{14-177}$ (2.9 mg) show that these constructs provide a greater protein yield compared to the original construct $TcGPXI_{9-177}$ (1.08 mg) and construct $TcGPXI_{18-177}$ (0.81 mg). It was observed through routine purification procedures that there was protein loss through non-specific binding to Ni-NTA resin, of which efforts were made to minimise.

In order to optimise protein expression, the *E. coli* expression cell lines Rosetta(DE3)pLysS (Novagen) and BL21(DE3), C41 and C43 (manufactured in-house competent cells, routinely used in our laboratory) were investigated. Following transformation of plasmid constructs for variants $TcGPXI_{13-177}$, $TcGPXI_{14-177}$ and the original template $TcGPXI_{9-177}$ into the various bacterial strains, protein was purified from the expression of 0.5 litre cell cultures of each host strain harbouring TcGPXI plasmid constructs. The resultant proteins were analyses using SDS-PAGE analysis (Figures 3.5-3.6) and Table 3.2 shows a comparison of the various His-tagged recombinant protein yields obtained. The results of the protein expression test show that overall $TcGPXI_{14-177}$ and $TcGPXI_{13-177}$ provided greater protein yields compared to protein purified from the $TcGPXI_{9-177}$ construct. The *E. coli* host bacterial strain Rosetta(DE3)pLysS harbouring the plasmid construct $TcGPXI_{14-177}$ was used to carry out structural studies and herein designated wild type TcGPXI for this study.

In order to produce sufficient protein for structural studies e.g. crystallisation trials and nuclear magnetic resonance spectroscopy (NMR), two litre cell cultures of *E. coli* Rosetta(DE3)pLysS harbouring the wild type TcGPXI plasmid construct was grown and the protein was expressed and purified as previously described. Figure 3.7 shows SDS-PAGE analysis of the wild type TcGPXI fractions of His-tagged recombinant protein, Ni-NTA resin, wash fractions and cleaved wild type TcGPXI from purification. The band at approximately 18 kDa indicates the presence of wild type TcGPXI protein following removal of the N-terminal histidine tag used for purification. Wild type TcGPXI consistently yielded approximately 10 mg of cleaved protein from two litres of cell culture following purification with immobilised metal affinity chromatography. The wild type TcGPXI protein was extensively dialysed into appropriate buffer in the presence or absence of reducing agent depending on

Sample	DH5α	BL21(DE3)	Rosetta(DE3)pLysS	C41	C43
$TcGPXI_{9-177}$	1.08 mg	3.24 mg	2.5 mg	-	-
$TcGPXI_{13-177}$	3.2 mg	4.43 mg	7.1 mg	3.5 mg	4.8 mg
$TcGPXI_{14-177}$	2.9 mg	6.0 mg	7.1 mg	4.4 mg	4.4 mg

Table 3.2: Comparison of protein amounts from various *E. coli* bacterial strains

Amounts of His-tagged recombinant protein obtained from 0.5 L test expression for $TcGPXI_{9-177}$ (template construct from BCSB), $TcGPXI_{13-177}$ and $TcGPXI_{14-177}$ in *E. coli* strains (DH5α, BL21(DE3), Rosetta(DE3)pLysS, C41 and C43).

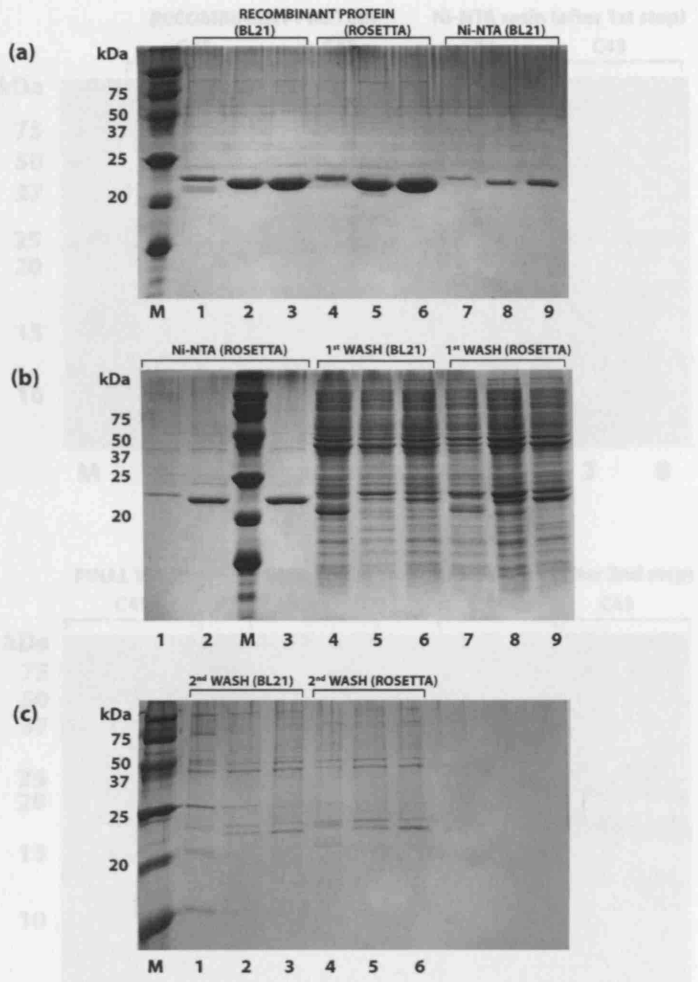


Figure 3.5: Test expression of TcGPXI constructs in various *E. coli* strains

SDS-PAGE analysis of purification of constructs $TcGPXI_{9\text{-}177}$, $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$, from expression in 0.5 L *E. coli* Rosetta(DE3)pLysS and BL21(DE3) cell cultures

(a) His-tagged protein (~ 22kDa) of $TcGPXI_{9\text{-}177}$, $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$ constructs from BL21(DE3) (lanes 1-3 respectively) and Rosetta(DE3)pLysS (lanes 4-6 respectively) cell lines; Ni-NTA resin from purification (after elution) from BL21(DE3) for constructs $TcGPXI_{9\text{-}177}$, $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$ (lanes 7-9 respectively).

(b) NiNTA resin of $TcGPXI_{9\text{-}177}$, $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$ from purification from Rosetta(DE3)pLysS (lanes 1-3 respectively). First wash fractions for $TcGPXI_{9\text{-}177}$, $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$: BL21(DE3) (lanes 4-6 respectively); Rosetta(DE3)pLysS (lanes 7-9 respectively).

(c) Second wash fractions for $TcGPXI_{9\text{-}177}$, $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$: BL21(DE3) (lanes 1-3 respectively); Rosetta(DE3)pLysS (lanes 4-6 respectively).

Molecular weight marker is denoted (M). Cells were treated equally through purification and equal sample volumes were loaded onto the gel.

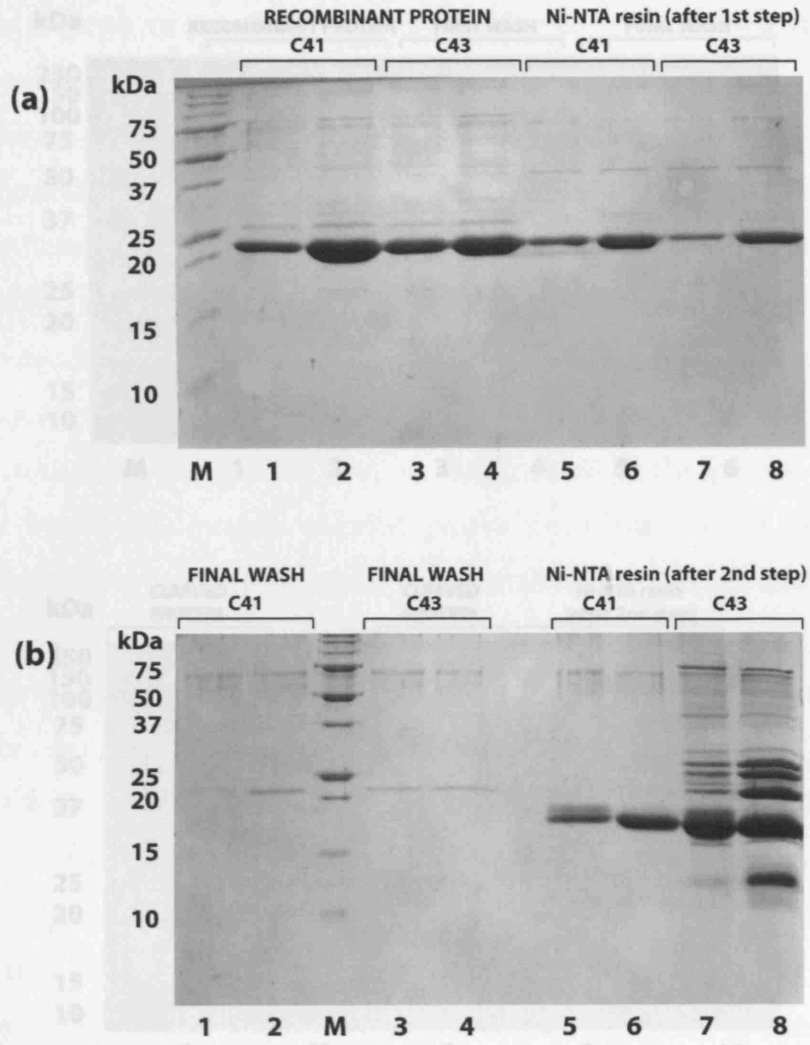


Figure 3.6: Test expression of TcGPXI constructs in *E. coli* C41 and C43 strains

SDS-PAGE analysis of purification of constructs $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$, from expression in 0.5 L *E. coli* C41 and C43 cell cultures. Cells were treated equally through purification and equal sample volumes were loaded onto the gels.

(a) His-tagged protein (~ 22kDa) of $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$ constructs from C41 (lanes 1-2 respectively) and C43 (lanes 3-4 respectively) cell lines; Ni-NTA resin from purification for constructs $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$: strain C41 (lanes 5-6 respectively) and C43 (lanes 7-8 respectively).

(b) Final wash fractions for $TcGPXI_{13\text{-}177}$, $TcGPXI_{14\text{-}177}$: C41 cell line (lanes 1-2 respectively) and C43 cell line (lanes 3-4); cleaved $TcGPXI_{13\text{-}177}$ and $TcGPXI_{14\text{-}177}$ proteins (lanes 5-6 respectively); Ni-NTA resin from removal of TEV protease, affinity tag and uncleaved protein for $TcGPXI_{13\text{-}177}$ and $TcGPXI_{14\text{-}177}$ (lanes 7-8 respectively).

Molecular weight marker denoted (M).

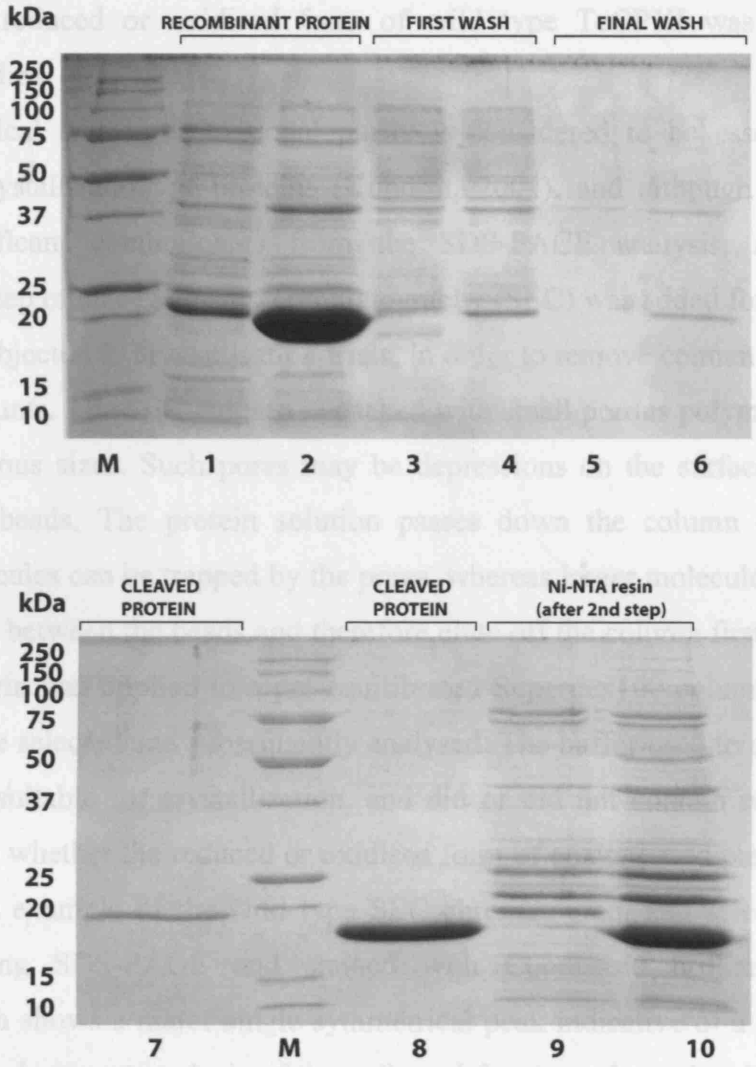


Figure 3.7: Wild type TcGPXI ($TcGPXI_{14-177}$) expression in Rosetta(DE3)pLysS host strain

SDS-PAGE analysis of protein expression of selected new wild type TcGPXI construct ($TcGPXI_{14-177}$) from 2 L of host strain *E. coli* Rosetta(DE3)pLysS compared to original template construct $TcGPXI_{9-177}$. His-tagged recombinant $TcGPXI_{9-177}$ (0.55 mg/ml) and $TcGPXI_{14-177}$ (1.8 mg/ml) eluted proteins (lanes 1-2 respectively); first wash fractions for $TcGPXI_{9-177}$ and $TcGPXI_{14-177}$ (lanes 3-4 respectively); final wash fractions for $TcGPXI_{9-177}$ and $TcGPXI_{14-177}$ (lanes 5-6 respectively); cleaved (His-tag removed) proteins for $TcGPXI_{9-177}$ (0.14 mg/ml) and $TcGPXI_{14-177}$ (0.92 mg/ml) (lanes 7-8 respectively); Ni-NTA resin from removal of TEV protease, affinity tag and uncleaved protein for $TcGPXI_{9-177}$ and $TcGPXI_{14-177}$ (lanes 9-10 respectively). Molecular weight marker denoted (M). Samples were treated equally and equal volumes were loaded onto the gel.

whether the reduced or oxidised form of wild type TcGPXI was required for structural studies.

Chemical and conformational purity is considered to be essential for the successful crystallisation of proteins (Kundrot, 2004), and although there are no visible significant contaminants from the SDS-PAGE analysis, an additional purification step of size exclusion chromatography (SEC) was added for proteins that were to be subjected to crystallisation trials, in order to remove contaminants present in small amounts. The SEC column is packed with small porous polymer beads with pores of various sizes. Such pores may be depressions on the surface or channels through the beads. The protein solution passes down the column during which smaller molecules can be trapped by the pores, whereas larger molecules simply pass more quickly between the beads and therefore elute off the column first. The cleaved purified protein was applied to a pre-equilibrated Superdex100 column from which fractions were selected and subsequently analysed. The buffer used to equilibrate the column was suitable for crystallisation, and did or did not contain reducing agent depending on whether the reduced or oxidised form of enzyme was prepared. Figure 3.8 shows an example of the wild type SEC chromatogram and selected fractions analysed using SDS-PAGE and stained with Coomassie brilliant blue. The chromatogram shows a major single symmetrical peak indicative of a single protein species, and subsequent analysis of the collected fractions shows bands representing wild type TcGPXI at approximately 18 kDa without any observable protein degradation. The purified fractions were then pooled. Overall pure and stable wild type TcGPXI was consistently produced and concentrated to a final concentration of 15-20 mg/ml. Wild type TcGPXI protein was used to carry out crystallisation trials, NMR spectroscopy and kinetic investigations (discussed in Chapter 4).

3.4 TcGPXII, TbPxI and TbPxII construct protein expression

Expression vectors containing DNA constructs coding for TbPxI, TbPxII, TcGPXIIa and TcGPXIIb were obtained. The respective plasmids were transformed into *E. coli* Top10 and Rosetta(DE3)pLysS host strains and 100 ml cell cultures were grown in Luria-Bertani (LB) broth to an OD_{600} of 0.6-0.8 at 37 °C. Protein was expressed by induction with IPTG to a final concentration of 1mM, followed by 3-4

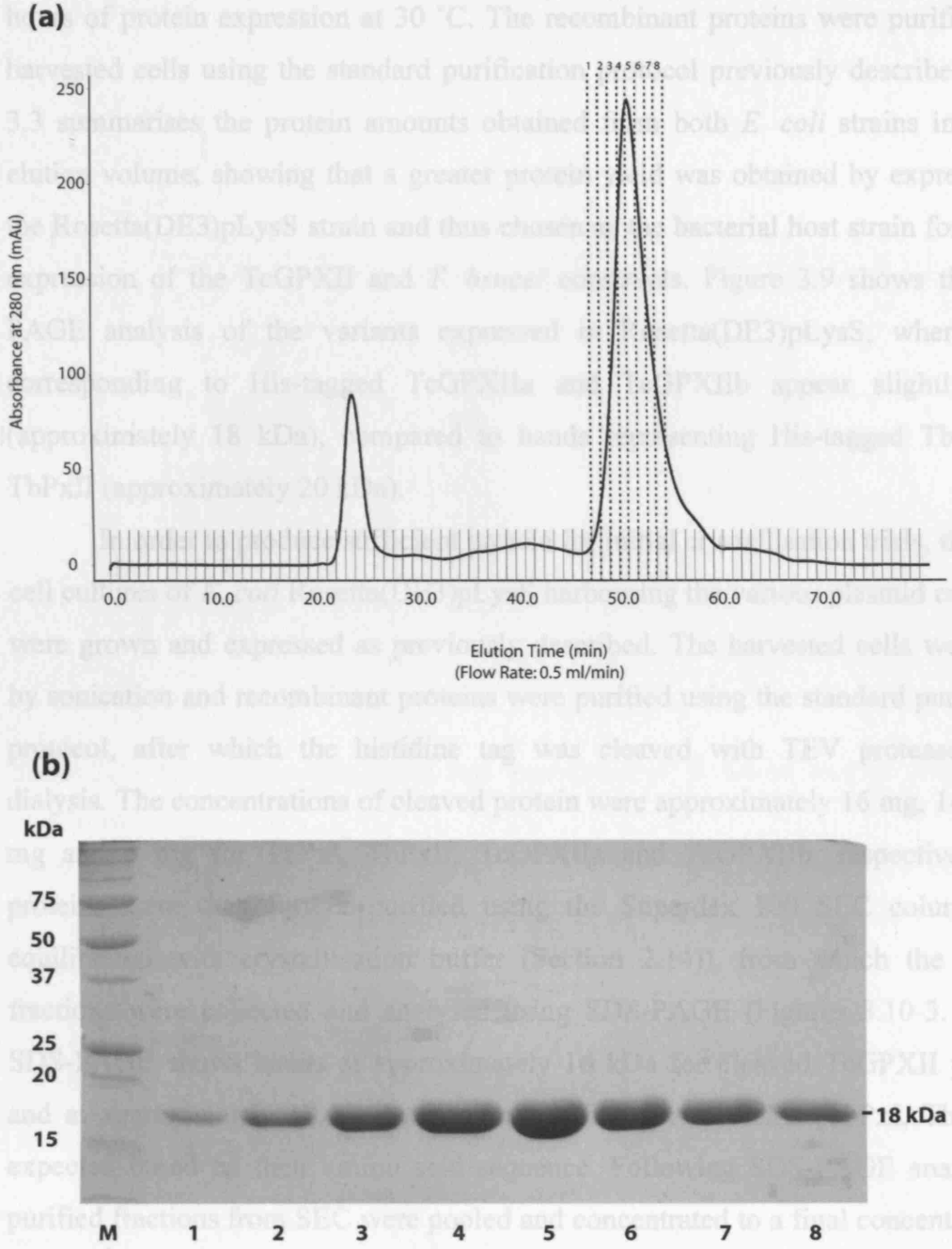


Figure 3.8: Size exclusion chromatography (SEC) of wild type TcGPXI ($TcGPXI_{14-177}$) protein

(a) The SEC chromatogram (pre-equilibrated and calibrated) of cleaved wild type TcGPXI (affinity tag removed). The chromatogram shows smaller peak (void volume) and majority single peak (wild type TcGPXI) coherent with the presence of a single protein species. Fractions 1-8 were collected (highlighted by dashed lines) and analysed by SDS-PAGE (b) showing purified cleaved wild type TcGPXI protein at approximately 18 kDa. Molecular marker is denoted (M). Equal sample volumes were loaded onto the gel.

hours of protein expression at 30 °C. The recombinant proteins were purified from harvested cells using the standard purification protocol previously described. Table 3.3 summarises the protein amounts obtained from both *E. coli* strains in a 3 ml elution volume, showing that a greater protein yield was obtained by expression in the Rosetta(DE3)pLysS strain and thus chosen as the bacterial host strain for protein expression of the TcGPXII and *T. brucei* constructs. Figure 3.9 shows the SDS-PAGE analysis of the variants expressed in Rosetta(DE3)pLysS, where bands corresponding to His-tagged TcGPXIIa and TcGPXIIb appear slightly lower (approximately 18 kDa), compared to bands representing His-tagged TbPxI and TbPxII (approximately 20 kDa).

In order to produce sufficient protein for initial crystallisation trials, three litre cell cultures of *E. coli* Rosetta(DE3)pLysS harbouring the various plasmid constructs were grown and expressed as previously described. The harvested cells were lysed by sonication and recombinant proteins were purified using the standard purification protocol, after which the histidine tag was cleaved with TEV protease during dialysis. The concentrations of cleaved protein were approximately 16 mg, 14 mg, 13 mg and 6 mg for TbPxI, TbPxII, TcGPXIIa and TcGPXIIb, respectively. The proteins were then further purified using the Superdex 100 SEC column (pre-equilibrated with crystallisation buffer (Section 2.14)), from which the purified fractions were collected and analysed using SDS-PAGE (Figures 3.10-3.13). The SDS-PAGE shows bands at approximately 16 kDa for cleaved TcGPXII proteins, and at approximately 18.5 kDa for cleaved *T. brucei* proteins (TbPxI, TbPxII) as expected based on their amino acid sequence. Following SDS-PAGE analysis the purified fractions from SEC were pooled and concentrated to a final concentration of approximately 15 mg/ml and used for initial crystallisation trials.

Sample	*E. coli* Top10	*E. coli* Rosetta(DE3)pLysS
TbPxI	0.19 mg	2.1 mg
TbPxII	0.82 mg	1.28 mg
TcGPXIIa	0.48 mg	0.8 mg
TcGPXIIb	0.56 mg	0.8 mg

Table 3.3: Protein amounts from test expression of *T. brucei* and *T. cruzi* GPXII constructs

Comparison of protein amounts (mg) from purification of constructs TbPxI, TbPxII, TcGPXIIa and TcGPXIIb expressed in 100 ml cell cultures of *E. coli* strains Top10 and Rosetta(DE3)pLysS.

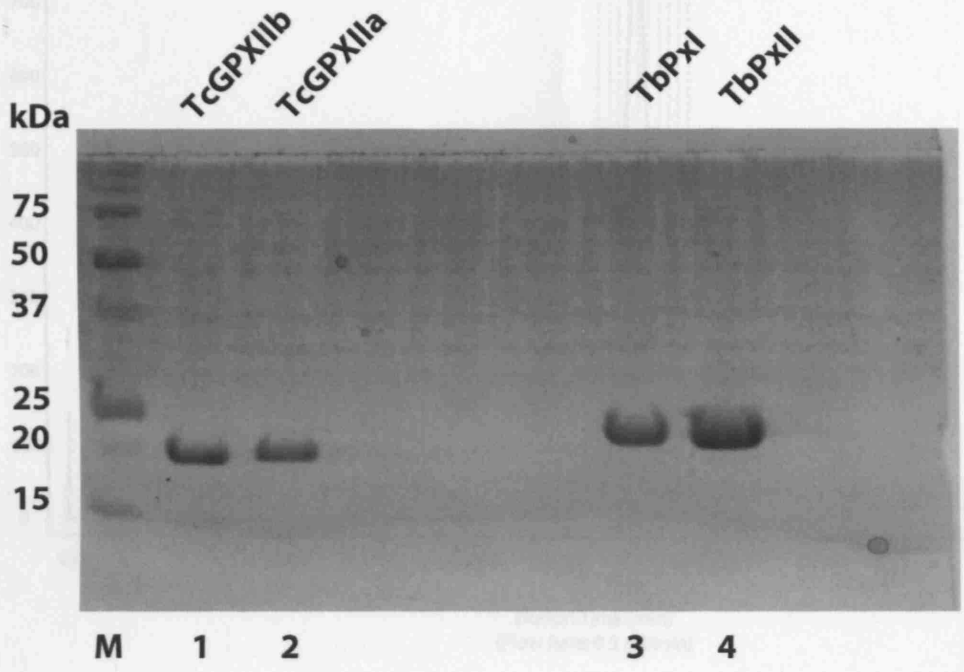


Figure 3.9: Test expression of *T. brucei* and *T. cruzi* GPXII constructs

SDS-PAGE analysis of purified His-tagged proteins from expression in 100 ml Rosetta(DE3)pLysS host strain: TcGPXIIb $_{5\text{-}160}$ (lane 1), TcGPXIIa $_{69\text{-}225}$ (lane 2), $TbPxI_{1\text{-}166}$ (lane3) and $TbPxII_{5\text{-}169}$ (lane 4). Molecular weight marker denoted (M). Samples were treated equally through purification and equal sample volumes were loaded onto the gel.

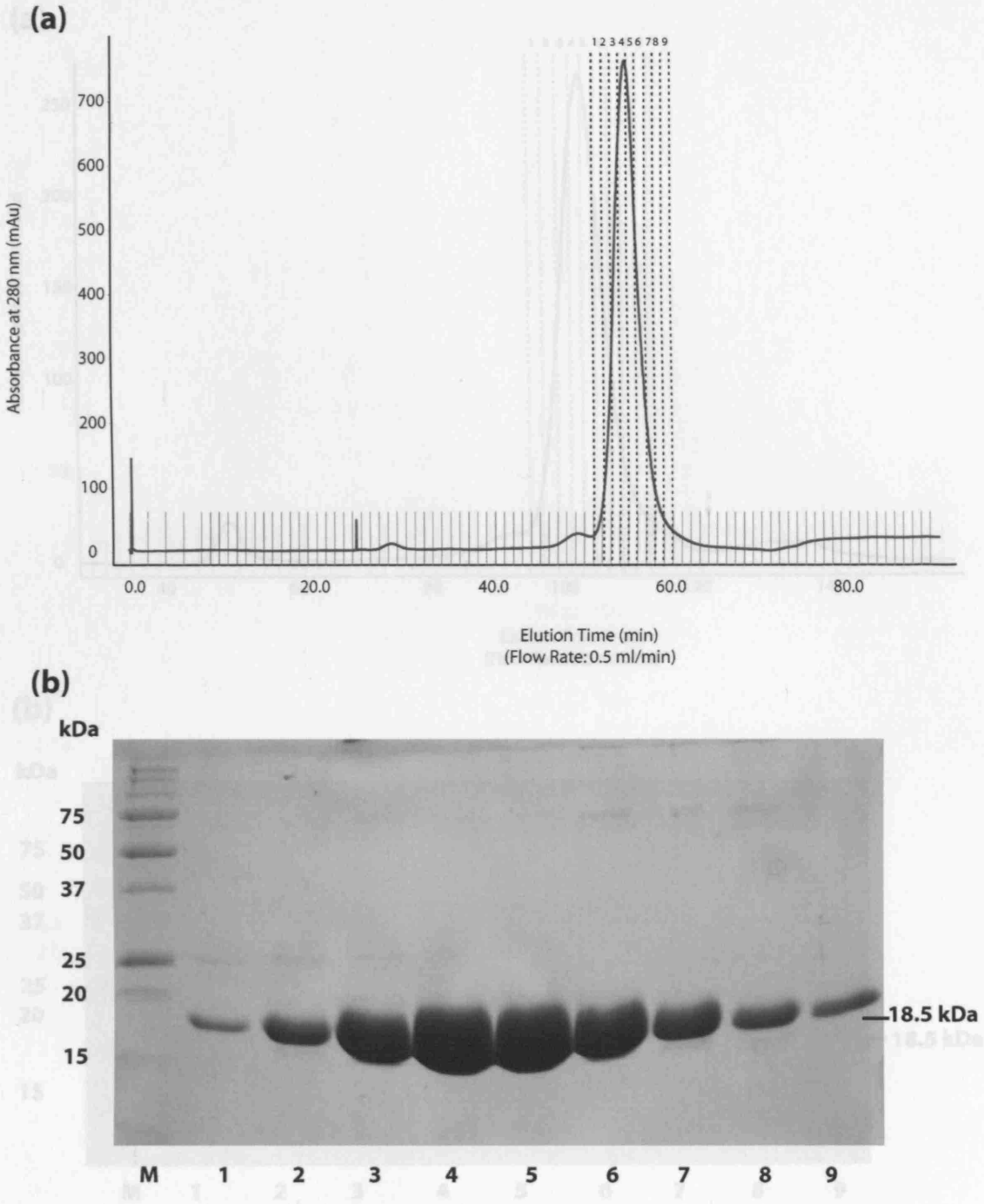


Figure 3.10: Size exclusion chromatography of *T. brucei* PxI construct ($TbPxI_{1-166}$)

(a) SEC chromatogram from purification of cleaved TbPxI protein (affinity tag removed) showing a single majority peak consistent with the presence of a single protein species. Fractions 1-9 indicated were collected and analysed using SDS-PAGE (b) showing bands corresponding to the resultant purified protein at approximately 18.5 kDa. Molecular weight marker is denoted (M). Equal sample volumes were loaded onto the gel.

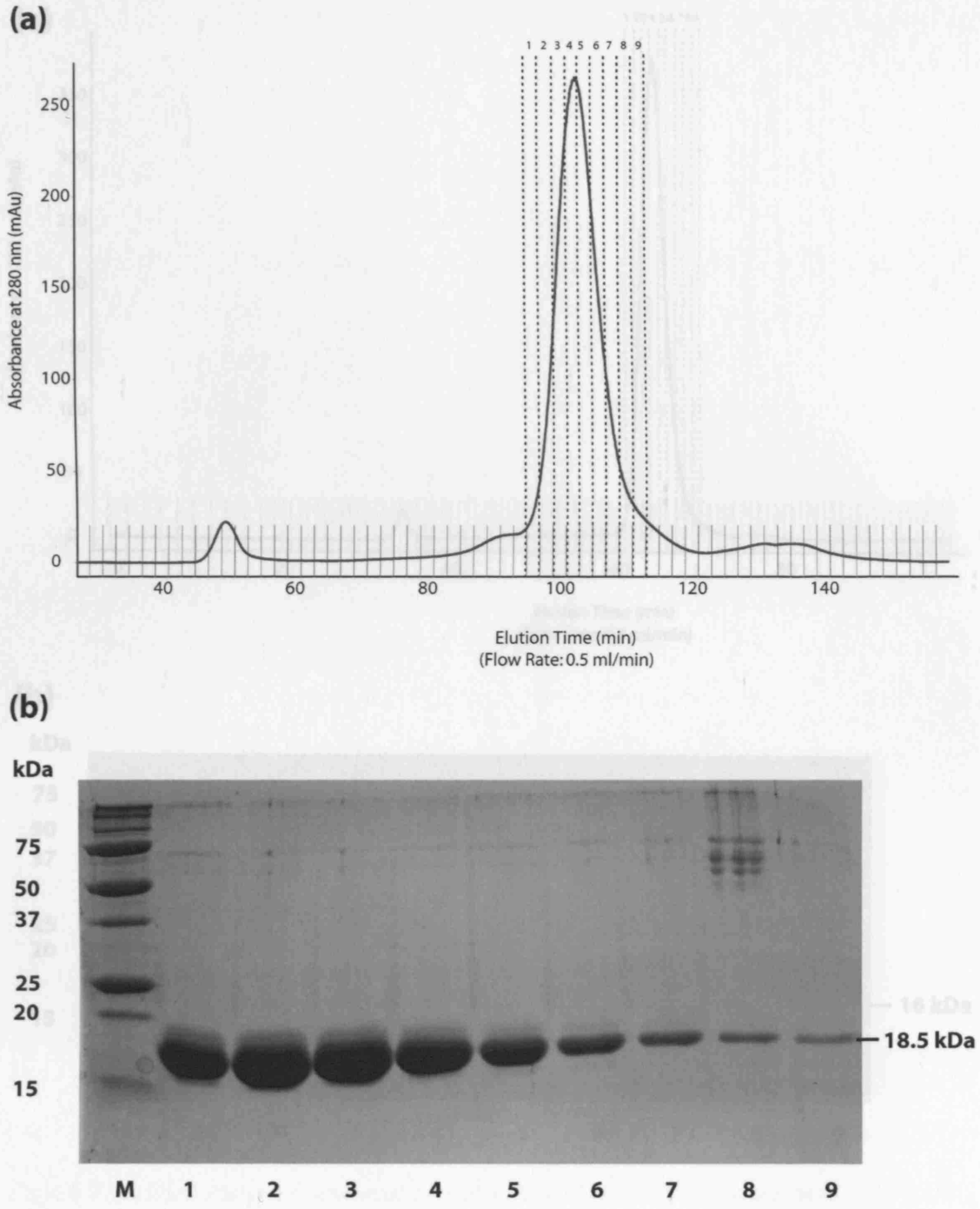


Figure 3.11: Size exclusion chromatography of *T. brucei* PxII construct ($TbPxII_{5-169}$)

(a) SEC chromatogram from purification of cleaved TbPxII protein (affinity tag removed) showing a single majority peak consistent with the presence of a single protein species. Fractions 1-9 indicated were collected and analysed using SDS-PAGE (b) showing bands corresponding to the resultant purified protein at approximately 18.5 kDa. Molecular weight marker is denoted (M). Equal sample volumes were loaded onto the gel.

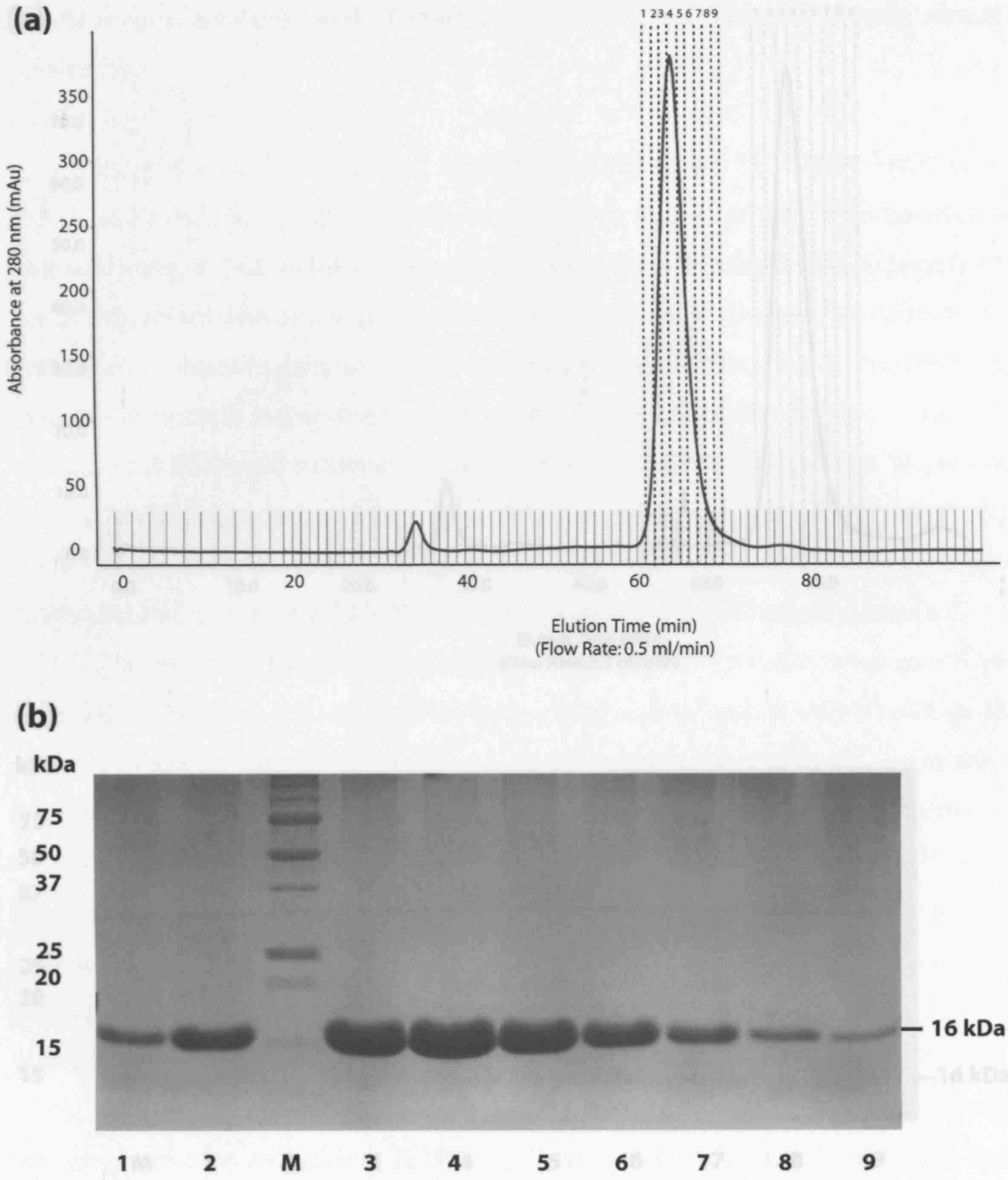


Figure 3.12: Size exclusion chromatography of *T.cruzi* GPXIIa construct (TcGPxIIa $_{69-225}$)

(a) SEC chromatogram from purification of cleaved TcGPXIIa protein (affinity tag removed) showing a single majority peak consistent with the presence of a single protein species. Fractions 1-9 indicated were collected and analysed using SDS-PAGE (b) showing bands corresponding to the resultant purified protein at approximately 16 kDa. Molecular weight marker is denoted (M). Equal sample volumes were loaded onto the gel.

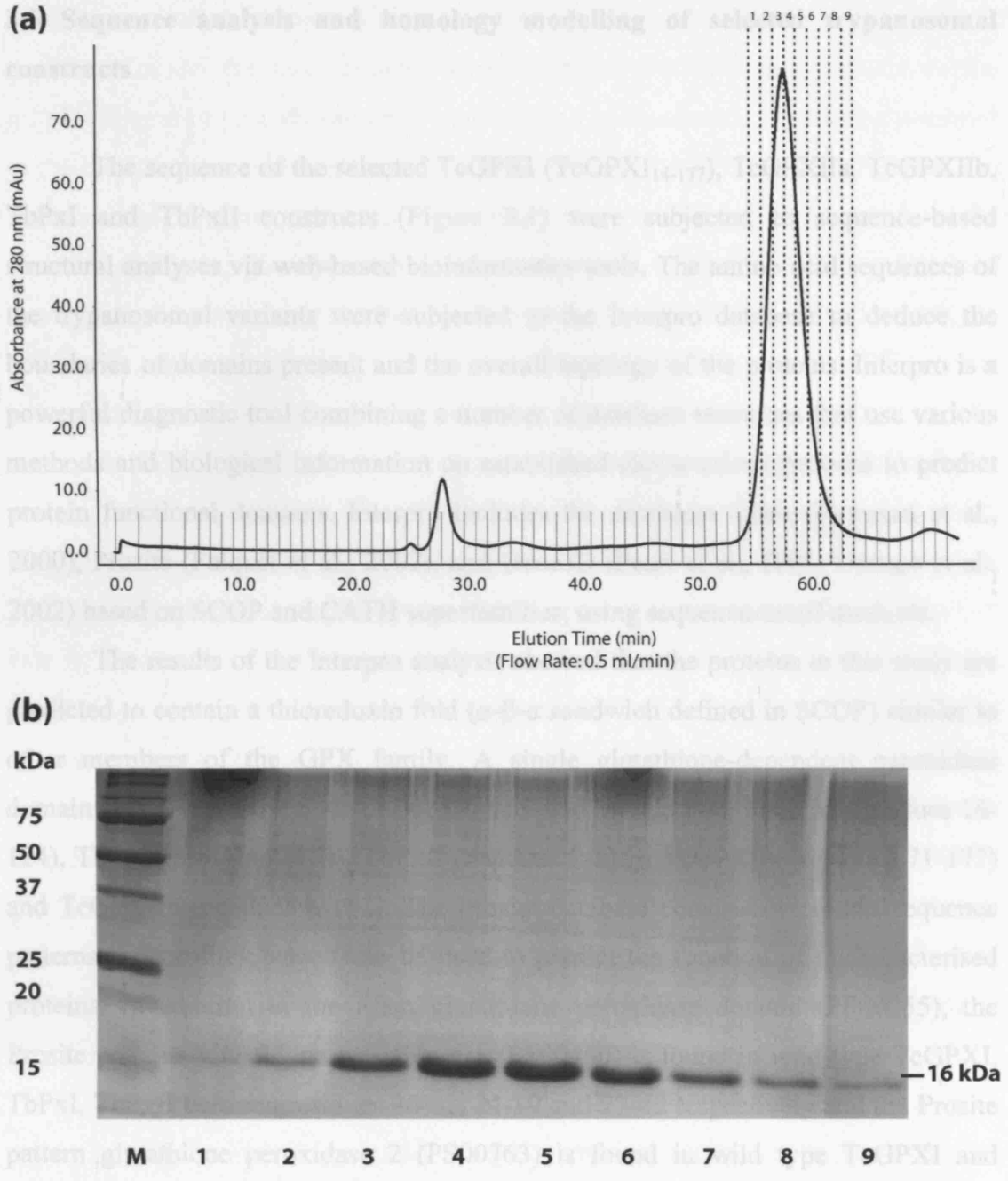


Figure 3.13: Size exclusion chromatography of *T.cruzi* GPXIIb construct (TcGPxIIb $_{5-160}$)

(a) SEC chromatogram from purification of cleaved TcGPXIIb protein (affinity tag removed) showing a single majority peak consistent with the presence of a single protein species. Fractions 1-9 indicated were collected and analysed using SDS-PAGE (b) showing bands corresponding to the resultant purified protein at approximately 16 kDa. Molecular weight marker is denoted (M). Equal sample volumes were loaded onto the gel.

3.5 Sequence analysis and homology modelling of selected trypanosomal constructs

The sequence of the selected TcGPXI ($TcGPXI_{14-177}$), TcGPXIIa, TcGPXIIb, TbPxI and TbPxII constructs (Figure 3.1) were subjected to sequence-based structural analyses via web-based bioinformatics tools. The amino acid sequences of the trypanosomal variants were subjected to the Interpro database to deduce the boundaries of domains present and the overall topology of the proteins. Interpro is a powerful diagnostic tool combining a number of database resources that use various methods and biological information on established characterised proteins to predict protein functional domains. Interpro includes the databases Pfam (Bateman et al., 2000), Prosite (Falquet et al., 2002), and Gene3D (Pearl et al., 2005; Orengo et al., 2002) based on SCOP and CATH superfamilies, using sequence-motif methods.

The results of the Interpro analysis showed that the proteins in this study are predicted to contain a thioredoxin fold (α-β-α sandwich defined in SCOP) similar to other members of the GPX family. A single glutathione-dependent peroxidase domain (Pfam accession number 00255), is found in wild type TcGPXI (residues 16-124), TbPxI (residues 4-112), TbPxII (residues 7-115), TcGPXIIa (residues 71-177) and TcGPXIIb (residues 6-112). The Prosite database consists of protein sequence patterns and profiles, which can be used to predict the function of uncharacterised proteins. In addition to the Pfam glutathione peroxidase domain (PF00255), the Prosite patterns glutathione peroxidase 1 (PS00460) is found in wild type TcGPXI, TbPxI, TbPxII between residues 36-51, 24-39 and 27-42 respectively and the Prosite pattern glutathione peroxidase 2 (PS00763) is found in wild type TcGPXI and TbPxII between residues 73-80 and 64-71 respectively. This is consistent with TcGPXI showing greater sequence homology to TbPxII (sequence identity 81%) than to TbPxI (sequence identity 76%). In contrast, the TcGPXII sequences have been shown to exhibit more similarity to other members of the GPX family than to trypanosomal sequences (Wilkinson et al., 2002b), as evident from the more distant sequence homology of TcGPXII to TcGPXI, TbPxI and TbPxII (sequence identity range 29-33%), compared to the sequence homology that TcGPXI shares with TbPxI and TbPxII (sequence identity of 76% and 81% respectively).

Homology modelling or comparative modelling involves the construction of an atomic-resolution three-dimensional model of a target of unknown structure based

on a known related structure as a template. In order to build the most accurate homology model the best template structure must first be found that shows the greatest degree of primary sequence conservation and secondary structure agreement to the target amino acid sequence. The most accurate models are obtained when the least number of amino acid residues are required to be inserted or deleted from the template sequence. The closest homologous template sequence of known structure can be found by searching the Protein Data Bank (PDB) database (Bernstein et al., 1977) using web-based search tools such as BLASTP (Altschul et al., 1990). It has been well-established that sequences showing sequence homology of as low as 33% sequence identity may be used to create models and infer the structural/functional features of a target sequence of experimentally undetermined tertiary structure (Sternberg and Zvelebil, 1990). After the template structure has been found an optimal multiple sequence alignment of the template and target is required, which can be generated using methods such as CLUSTALW (Larkin et al., 2007). However, close manual inspection is required of where insertions and deletions are placed as structural features are not considered in the CLUSTALW sequence alignment, and may therefore require manual editing to provide an optimally feasible alignment for modelling. Secondary structure assignment using the program DSSP (Kabsch and Sander, 1983) (based on the template structure) allows identification of secondary structural elements and facilitates insertions/deletions to be localised to structurally variable regions *e.g.* loop regions, as opposed to more ordered parts of the model.

The optimal sequence alignment of template and target sequence can then be used to generate a homology model using the software MODELLER (Marti-Renom et al., 2000; Eswar et al., 2007). MODELLER uses molecular dynamics to fit the target sequence to the template structure, based on the sequence alignment given whilst taking into consideration the conserved and variable regions. Restraints are applied to weight the structurally conserved regions more highly than the variable regions. MODELLER automatically performs energy minimisation during model building, to remove undesirable structural features such as bad contacts or non-optimal loop conformations and provide a model of the lowest energy conformation. Homology models can be validated by PROCHECK (Laskowski et al., 1993) which uses stereochemical parameters to evaluate the bond angles and bond lengths of every amino acid residue of the protein. It is important to remember that although

homology models are based on a template structure determined by experimental data, the output is a model based on another protein with least accuracy in structurally variable regions. The most accurate tertiary information for a target sequence can be obtained by experimental studies such as X-ray crystallography and NMR spectroscopy.

Homology models were created of wild type TcGPXI, TbPxI and TcGPXIIa. The closest homologues based on highest sequence identity to wild type TcGPXI and TbPxI are structures relating to the oxidised (PDB code: 2P5R) and reduced (PDB code: 2P5Q) forms of *Populus trichocarpa x deltoids* GPX5 (PtGPX5, accession number ABN59534), showing approximately 48 % sequence identity. The PtGPX5 template structures show a thioredoxin fold with the presence of a helix in the reduced form that undergoes conformational change upon oxidation facilitating the formation of an intramolecular disulphide bond. Figure 3.14a shows the multiple sequence alignment used to generate homology models (for both oxidised and reduced forms) for wild type TcGPXI and TbPxI enzymes and the PtGPX5 template sequence. The DSSP secondary structure prediction for the oxidised form is highlighted in red, and additional secondary structural elements (α4, α') associated with the reduced form are shown in blue. The catalytic residues (cysteine, glutamine and tryptophan) and regions of sequence conservation are highlighted.

The resultant homology models for wild type TcGPXI (Figure 3.15) and TbPxI (Figure 3.16) show a thioredoxin fold consisting of an α-β-α sandwich. The oxidised forms (Figure 3.15a and Figure 3.16a) show the presence of a disulphide bond otherwise absent in the reduced forms (Figure 3.15b and Figure 3.16b) shown to occur between analogous Cys48 and Cys96 in wild type TcGPXI; and between Cys36 and Cys84 in TbPxI.

Despite the helical region (α*; Figure 3.16a) assigned by DSSP in both conformations of the TbPxI homology models, the graphics program Pymol has not drawn this in the reduced TbPxI model (Figure 3.16b). Additionally, the absence of α' (small helix of only 3 residues) in the reduced form of the TbPxI model, may be explained by the fact that when building a homology model the program MODELLER will account for the amino acids that are present and carry out energy minimisation, thus producing the most feasible secondary structural elements. In a similar fashion Pymol (used for imaging the homology models) may contain its own

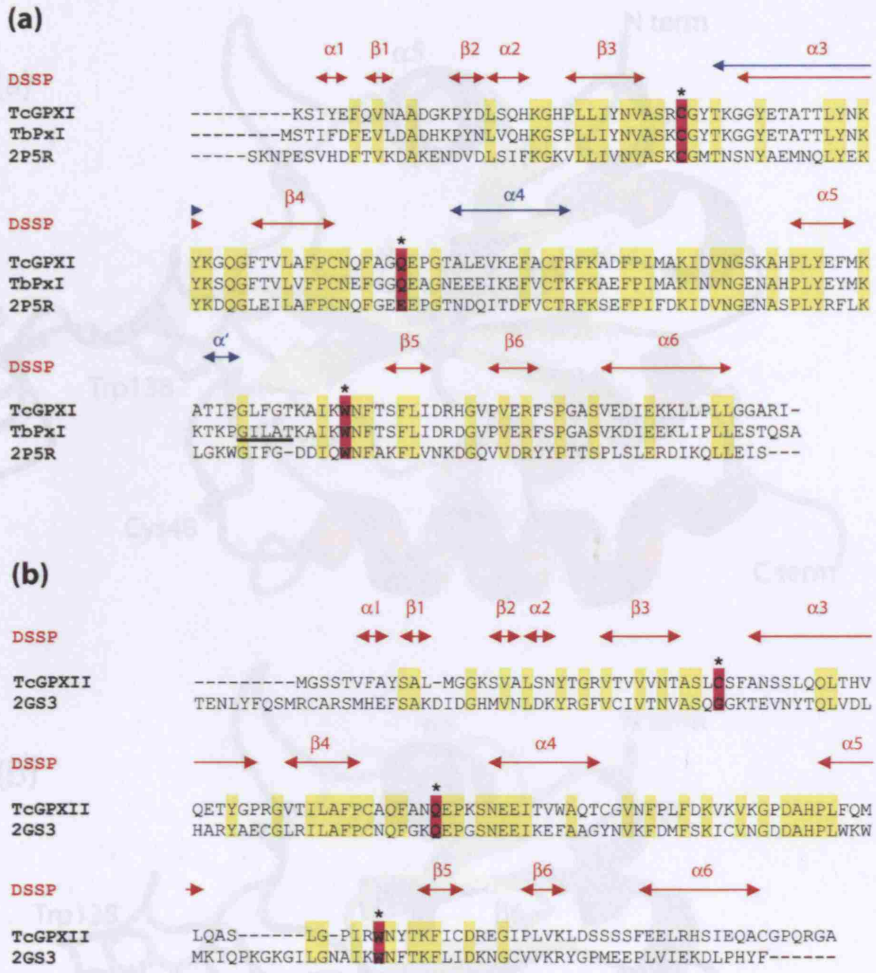


Figure 3.14: Multiple sequence alignments used to generate homology models

Regions of sequence conservation (yellow) and conserved catalytic triad of cysteine, glutamine and tryptophan (pink and *) are highlighted.

(a) Multiple sequence alignment of wild type TcGPXI, TbPxI and template structure of homologue PtGPX5 in the oxidised form (PDB code: 2P5R). Secondary structure assignment (DSSP) is shown in red, with additional secondary structural elements present in the crystal structure of reduced PtGPX5 (PDB code: 2P5Q) shown in blue. Sequence region giving rise to an additional helical structure found only in homology models of TbPxI is underlined.

(b) Multiple sequence alignment of TcGPXII and template structure of the homologue human GPX4 (PDB code: 2GS3). The secondary structure assignment (DSSP) is shown in red.

TcGPXI, wild type *T. cruzi* GPXI ($TcGPXI_{14-177}$); **TbPxI**, *T. brucei* PxI ($TbPxI_{1-166}$); **2P5R**, *P. trichocarpa x deltoides* GPX5 (accession no. ABN59534) template structure in the oxidised form; **TcGPXII**, *T. cruzi* GPXIIa ($TcGPXIIa_{69-225}$); **2GS3**, *H. sapiens* PHGPX4 (accession no. P36969).

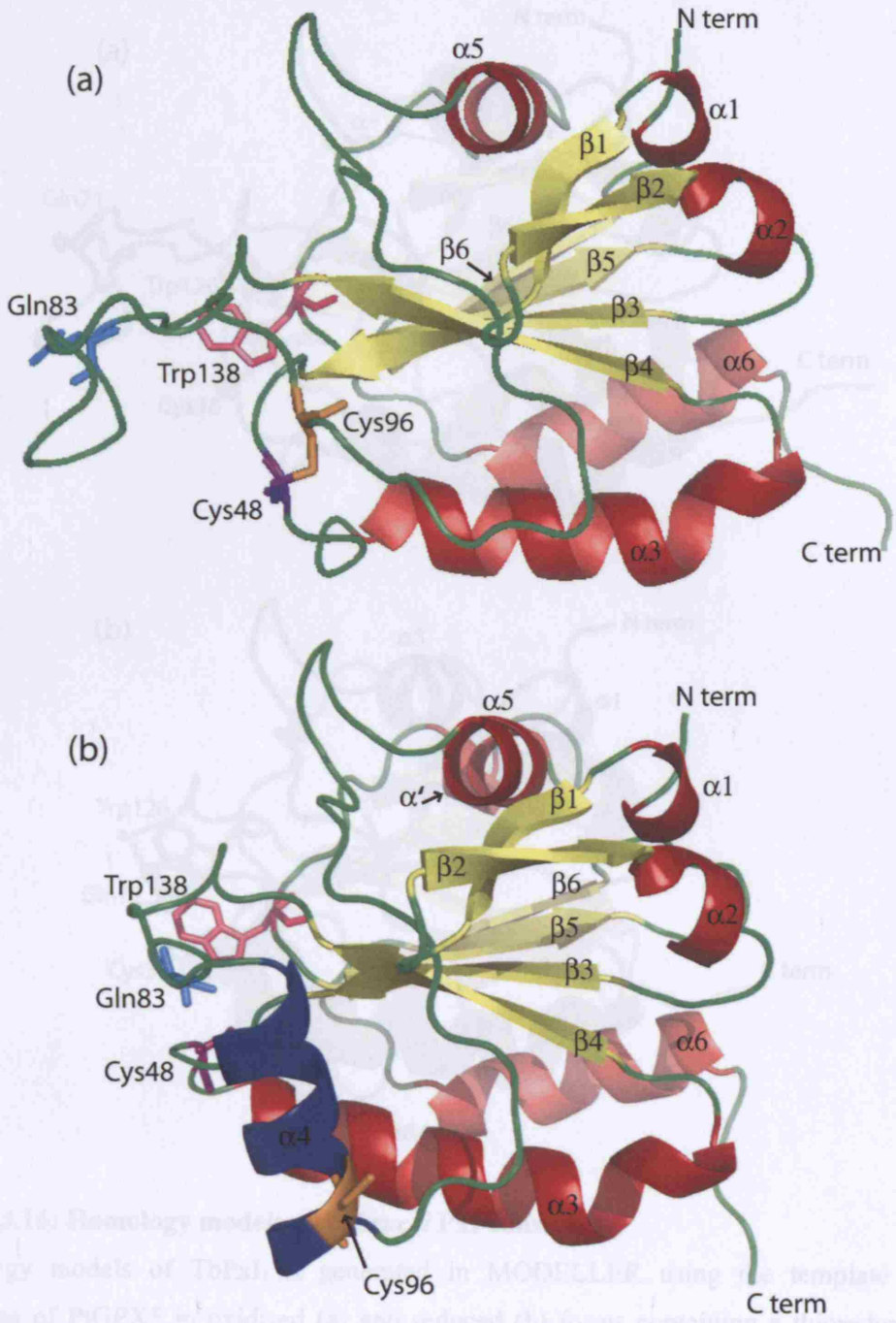


Figure 3.15: Homology models of wild type *T. cruzi* GPXI

Homology models of wild type $TcGPXI_{14-177}$ generated in MODELLER using the template crystal structures of PtGPX5 in oxidised (a) and reduced (b) forms comprising a thioredoxin fold of overall α-β-α topology. Catalytic triad of cysteine 48 (purple), glutamine 83 (cyan) and tryptophan 138 (pink) are illustrated. Additional resolving cysteine 96 (orange) participating in a disulphide bond with cysteine 48 in the oxidised form (a) is highlighted. The homology model of the reduced conformation contains the resolving cysteine (Cys96) localised in α4 (helix shown in blue). (Figures generated by Pymol (DeLano, W. L., 2002)).

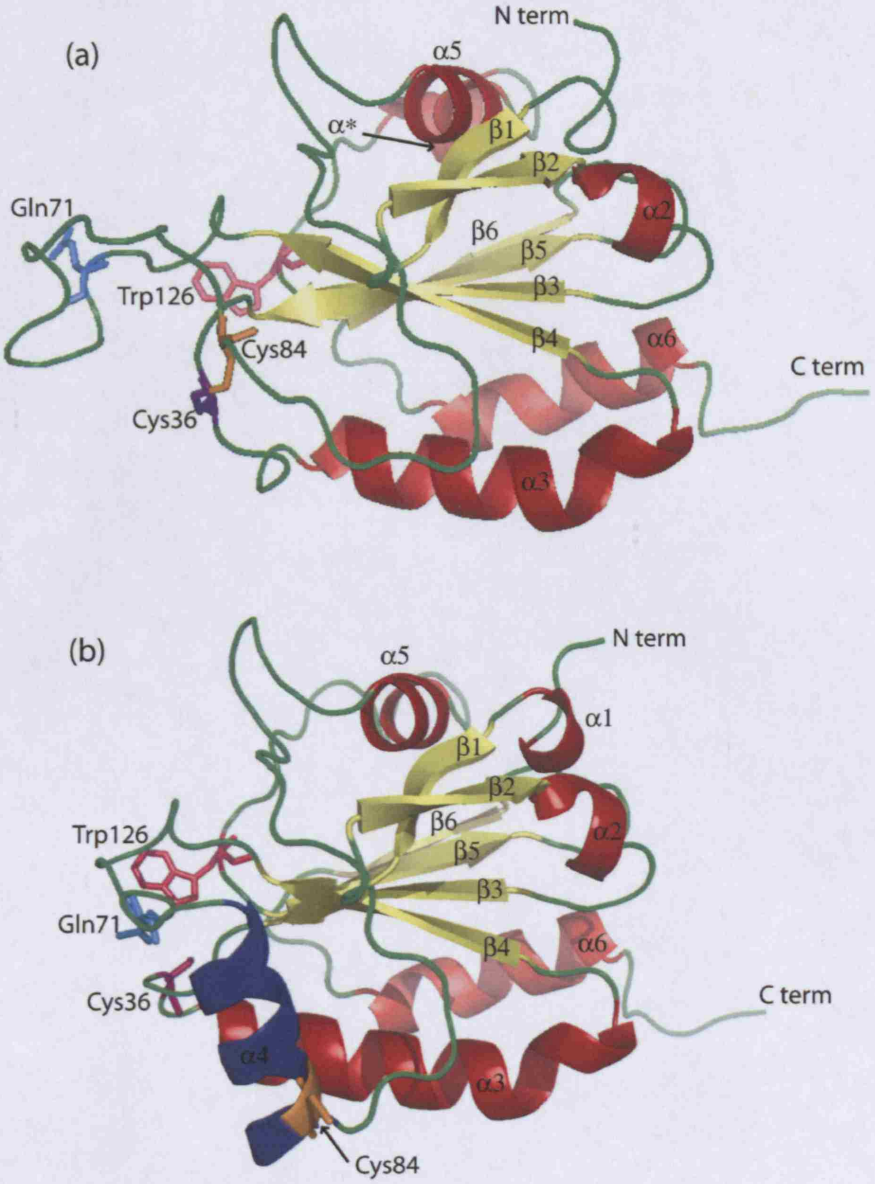


Figure 3.16: Homology models of *T. brucei* PxI construct

Homology models of $TbPxI_{1\text{-}166}$ generated in MODELLER using the template crystal structures of PtGPX5 in oxidised (a) and reduced (b) forms containing a thioredoxin fold comprising an overall α-β-α sandwich topology. Catalytic triad of cysteine 36 (purple), glutamine 71 (cyan) and tryptophan 126 (pink) are illustrated. Additional resolving cysteine 84 (orange) participating in a disulphide bond with cysteine 36 in the oxidised form (a) is highlighted. The homology model of the reduced form (b) shows additional helical region (α4) present only in the reduced form (α' does not occur in the model of TbPxI in the reduced form), where the resolving cysteine (Cys84) resides in α4 (helix shown in blue). Secondary structure assignment by DSSP indicates the presence of α1 and α* (sequence motif GILAT) in both models, although not drawn by Pymol in (a) and (b) respectively. (Figures generated by Pymol (DeLano, W. L., 2002)).

constraints dependent on factors such as the primary sequence and the propensity of certain amino acids to participate in secondary structural elements, therefore some secondary structural elements, despite assignment made by DSSP, are not drawn in the Pymol output image *e.g.* α1 and α* are not drawn in the oxidised and reduced forms of TbPxI respectively of Figure 3.16.

The closest homologue of known structure to TcGPXII is the crystal structure of human PHGPX4 (2GS3) showing 41% sequence identity. Figure 3.14b shows the multiple sequence alignment used to generate the homology model, where the secondary structure prediction (red) by DSSP is highlighted. The active site selenocysteine/cysteine, glutamine and tryptophan (crystal structure of 2GS3 contains a mutation of the active site selenocysteine to glycine) are highlighted in addition to sequence conservation. Figure 3.17 shows the resultant homology model exhibiting the expected thioredoxin fold consisting of an α-β-α sandwich, where despite their prediction by DSSP, the secondary structure elements β1 and β2 are not drawn by the program Pymol (DeLano, 2002).

From the resultant homology models it is shown that the amino acid sequences of the *T. cruzi* GPXI and *T. brucei* enzymes is able to be modelled in both the oxidised and reduced conformations based upon the template structures of PtGPX5. It may be suggested that these proteins show similar redox centres where it is postulated that *T. brucei* and TcGPXI sequences exhibit similar catalytic mechanisms. From the homology model of TcGPXII it may be suggested that this enzyme could display a reaction mechanism more similar to that of the classic human GPX enzymes rather than the atypical 2-Cys peroxiredoxin type mechanism employed by PtGPX5. Detailed mutational analysis and kinetic measurements in conjunction with structural examination of the redox centres would be required to elucidate the exact catalytic mechanisms and account for the variations in substrate specificities of these enzymes.

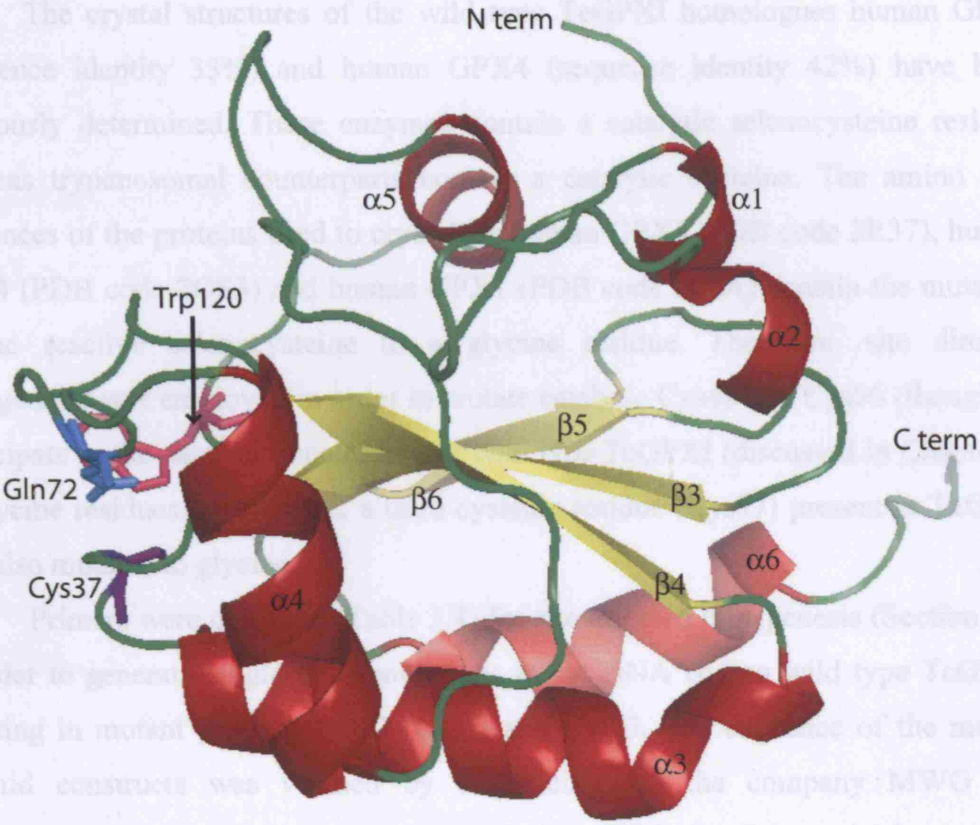


Figure 3.17: Homology model of *T. cruzi* GPXIIa construct

Homology model of TcGPXIIa $_{69\text{-}225}$ generated in MODELLER using a template crystal structure of human PHGPX4 (PDB code: 2GS3) that contains a thioredoxin fold consisting of an overall α-β-α sandwich topology. Catalytic triad of cysteine 37 (purple), glutamine 72 (cyan) and tryptophan 120 (pink) are illustrated. Secondary structural elements β1 and β2 (assigned by DSSP), are not drawn by Pymol. (Figures generated by Pymol (DeLano, W. L., 2002)).

3.6 Site directed mutagenesis of wild type TcGPXI

The crystal structures of the wild type TcGPXI homologues human GPX3 (sequence identity 33%) and human GPX4 (sequence identity 42%) have been previously determined. These enzymes contain a catalytic selenocysteine residue, whereas trypanosomal counterparts contain a catalytic cysteine. The amino acid sequences of the proteins used to crystallise human GPX3 (PDB code 2R37), human GPX4 (PDB code 2GS3) and human GPX1 (PDB code 2F8A) contain the mutation of the reactive selenocysteine to a glycine residue. Therefore, site directed mutagenesis was employed in order to mutate catalytic Cys48 and Cys96 (thought to participate in the catalytic mechanism of wild type TcGPXI (discussed in Chapter 4) to glycine residues. In addition, a third cysteine residue (Cys77) present in TcGPXI was also mutated to glycine.

Primers were designed (Table 3.4) for site directed mutagenesis (Section 2.6) in order to generate single point mutations in the DNA coding wild type TcGPXI, resulting in mutant proteins C48G, C77G and C96G. The sequence of the mutant plasmid constructs was verified by sequencing via the company MWG and successful clones transformed into *E. coli* Rosetta(DE3)pLysS bacterial host strain. In the same manner as wild type TcGPXI, three litre cell cultures were grown, expressed and the recombinant protein purified using Ni-NTA resin (Section 3.4). The affinity tag was removed by addition of TEV protease and the resultant cleaved protein analysed by SDS-PAGE (Figure 3.18). The purified, cleaved C48G, C96G and C77G mutant TcGPXI proteins were used to carry out NMR spectroscopy and kinetic measurements (Chapter 4). The mutant C48G and C96G TcGPXI proteins were subjected to further purification using SEC. Figures 3.19 and 3.20 show single symmetrical peaks of the SEC chromatographs coherent with the presence of a single protein species for wild type $TcGPXI_{14\text{-}177}$ mutants C48G and C96G. It was observed from the purification of C48G that there was some protein aggregation, which can be detected by the concentration of protein in the void volume compared to the concentration in the fraction peak of the eluted protein in Figure 3.19a.

The purified C48G and C96G TcGPXI mutant proteins were subjected to initial crystallisation studies in the absence of reducing agent.

$TcGPXI_{14-177}$ mutant		PRIMER
C48G	Forward	CAATGTGGCGAGTAGGGGTGGCTACACGAAGGGTG
	Reverse	CACCCTTCGTGTAGCCACCCCTACTCGCCACATTG
C77G	Forward	GTTCTGGCGTTCCCGGGTAACCAATTTGCCGGG
	Reverse	CCCGGCAAATTGGTTACCCGGGAACGCCAGAAC
C96G	Forward	GGTGAAGGAATTTGCCGGCACGCGGTTTAAGGCGG
	Reverse	CCGCCTTAAACCGCGTGCCGGCAAATTCCTTCACC

Table 3.4: Primers for site directed mutagenesis of wild type TcGPXI ($TcGPXI_{14-177}$)

Forward (5'→ 3') and reverse (3'→ 5') primer designs for site directed mutagenesis of wild type TcGPXI ($TcGPXI_{14-177}$). TcGPXI mutants were created (C48G, C77G, C96G) where residues Cys48, Cys77 and Cys96 were mutated to glycine.

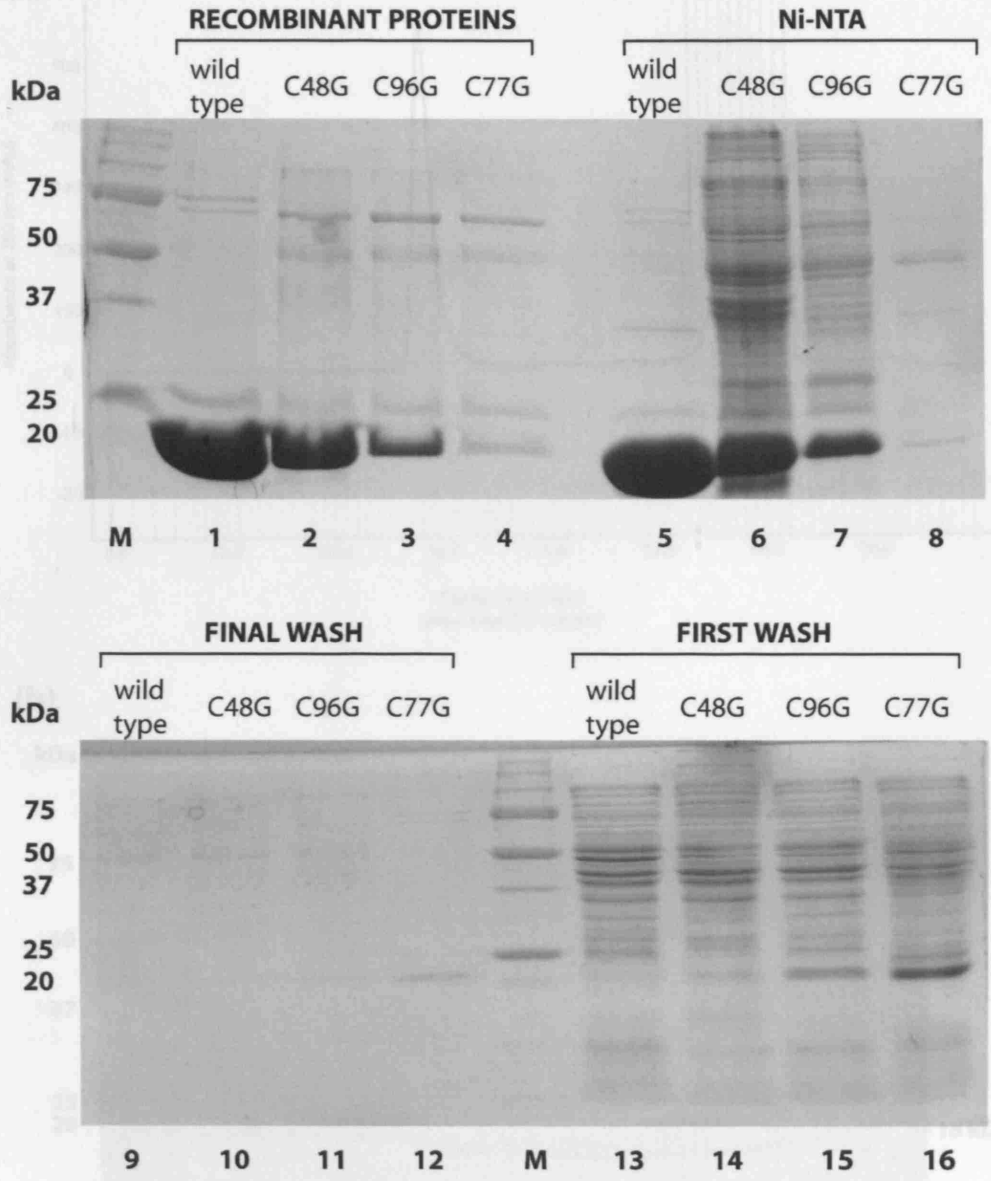


Figure 3.18: Site directed mutagenesis of wild type *T. cruzi* GPXI ($TcGPXI_{14-177}$)
SDS-PAGE analysis of purified C48G, C77G and C96G proteins from *E. coli* Rosetta(DE3)pLysS. Protein from wild type (2.6 mg/ml), C48G (2 mg/ml), C96G (1.05 mg/ml) and C77G (0.6 mg/ml) TcGPXI (lanes 1-4 respectively); Ni-NTA resin from wild type, C48G, C96G and C77G TcGPXI (lanes 5-8 respectively); final wash fractions for wild type, C48G, C96G and C77G TcGPXI (lanes 9-12 respectively); first wash fractions for wild type, C48G, C96G and C77G TcGPXI (lanes 13-16 respectively). Molecular weight marker is denoted (M). Samples were treated equally and equal volumes were loaded onto the gel.

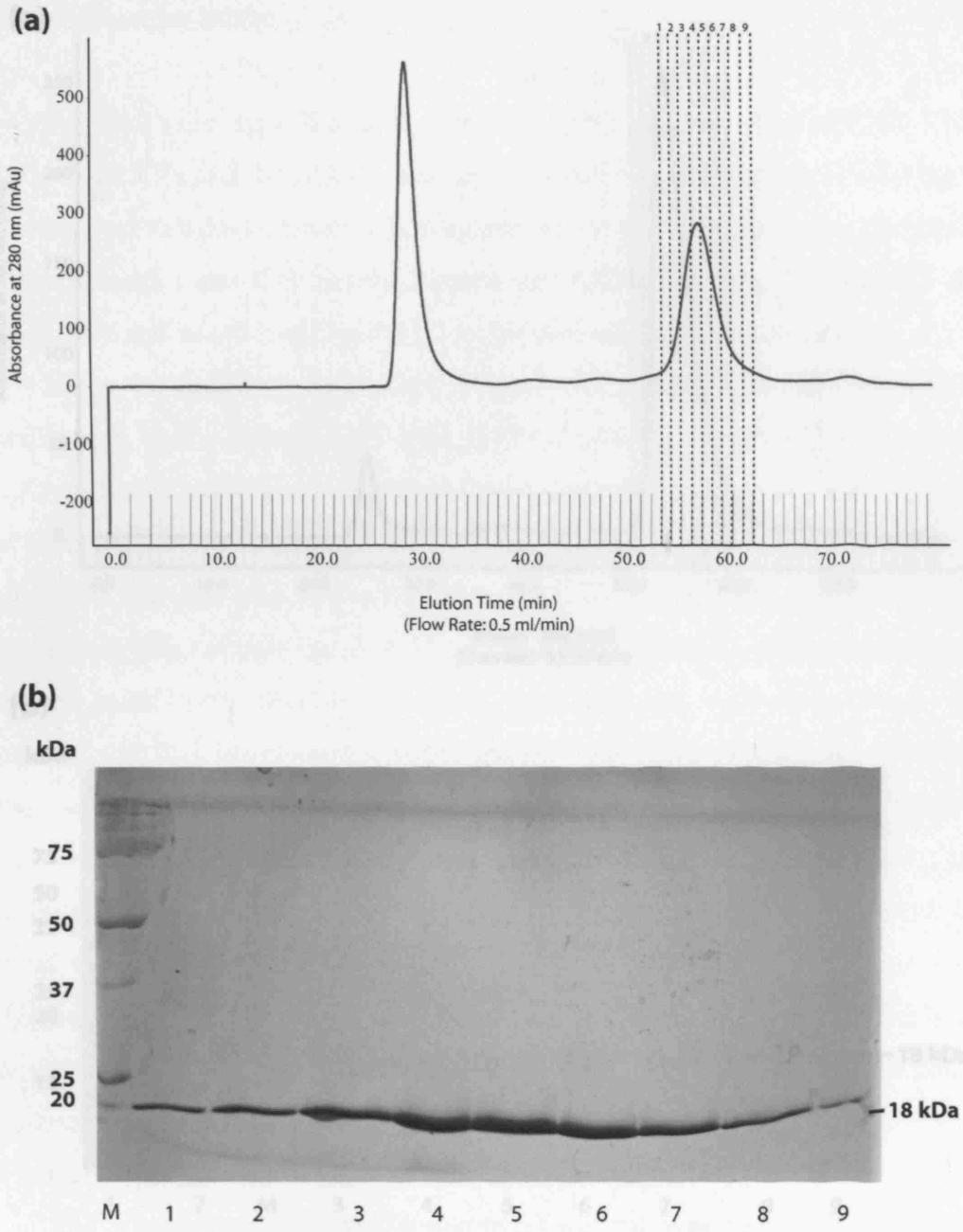


Figure 3.19: Size exclusion chromatography (SEC) of C48G mutant TcGPXI protein

(a) The SEC chromatogram of cleaved C48G mutant TcGPXI (affinity tag removed). The chromatogram shows a large peak (void volume) due to large amounts of contaminants in this preparation and smaller single peak (wild type TcGPXI) coherent with the presence of a single protein species. Fractions 1-9 of purified C48G mutant protein were collected and analysed by SDS-PAGE (b) showing purified cleaved C48G mutant TcGPXI protein at ~ 18 kDa. Molecular marker is denoted (M). Equal sample volumes were loaded onto the gel.

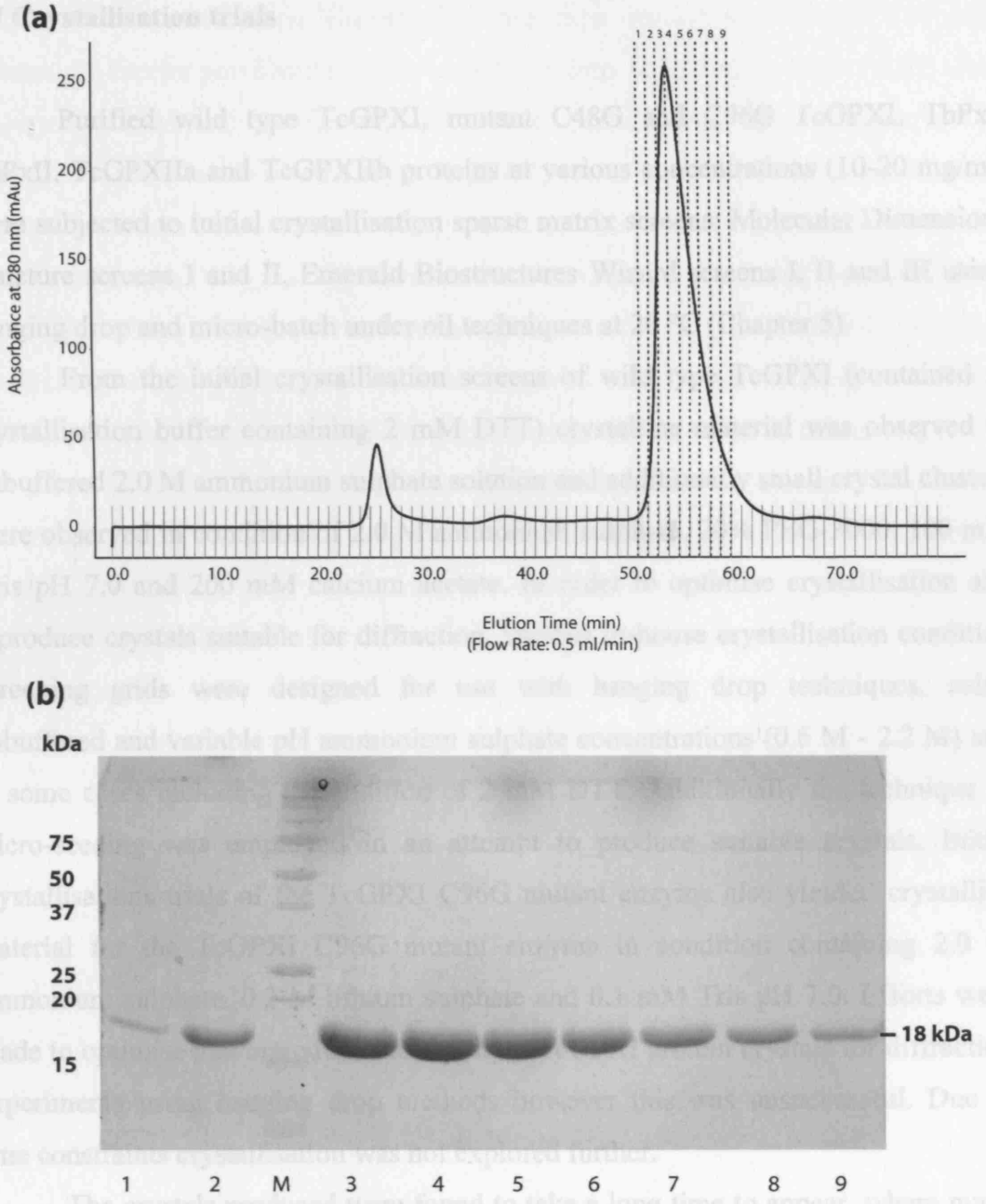


Figure 3.20: Size exclusion chromatography (SEC) of C96G mutant TcGPXI protein
(a) The SEC chromatogram of cleaved C96G mutant TcGPXI (affinity tag removed). The chromatogram shows smaller peak (void volume) and majority single peak (C96G mutant TcGPXI protein) coherent with the presence of a single protein species. Fractions 1-9 were collected and analysed by SDS-PAGE (b) showing purified cleaved C96G mutant TcGPXI protein at ~ 18 kDa. Molecular marker is denoted (M). Equal sample volumes were loaded onto the gel.

3.7 Crystallisation trials

Purified wild type TcGPXI, mutant C48G and C96G TcGPXI, TbPxI, TbPxII, TcGPXIIa and TcGPXIIb proteins at various concentrations (10-20 mg/ml) were subjected to initial crystallisation sparse matrix screens: Molecular Dimensions structure screens I and II, Emerald Biostructures Wizard screens I, II and III using hanging drop and micro-batch under oil techniques at 20 °C (Chapter 5).

From the initial crystallisation screens of wild type TcGPXI (contained in crystallisation buffer containing 2 mM DTT) crystalline material was observed in unbuffered 2.0 M ammonium sulphate solution and additionally small crystal clusters were observed in condition of 2.0 M ammonium sulphate, 20% PEG-3000, 100 mM Tris pH 7.0 and 200 mM calcium acetate. In order to optimise crystallisation and reproduce crystals suitable for diffraction, various in-house crystallisation condition screening grids were designed for use with hanging drop techniques, using unbuffered and variable pH ammonium sulphate concentrations (0.6 M - 2.2 M) and in some cases including the addition of 2 mM DTT. Additionally the technique of micro-seeding was employed in an attempt to produce suitable crystals. Initial crystallisations trials of the TcGPXI C96G mutant enzyme also yielded crystalline material for the TcGPXI C96G mutant enzyme in condition containing 2.0 M ammonium sulphate, 0.2 M lithium sulphate and 0.1 mM Tris pH 7.0. Efforts were made to optimise and reproduce these mutant TcGPXI protein crystals for diffraction experiments using hanging drop methods however this was unsuccessful. Due to time constraints crystallisation was not explored further.

The crystals produced were found to take a long time to appear, where many grew as clusters and it was difficult to control crystallisation conditions. It was found that crystals were observed only after a period of months, based on which it may be suggested that the reducing agent is no longer active therefore the wild type TcGPXI protein forms crystals in its oxidised state. The antioxidant cycle of wild type TcGPXI shows that upon reduction of the hydroperoxide substrate the enzyme itself becomes oxidised, upon which the reduced form is regenerated by interaction with the redox protein tryparedoxin. In order to obtain fully oxidised protein under controlled conditions and hence in an attempt to control crystallisation, a five-fold excess of t-butyl hydroperoxide (known TcGPXI substrate, (Wilkinson et al., 2000a)) was added to purified cleaved wild type TcGPXI protein and left at room

temperature for two hours. The solution was then applied to Superdex 100 SEC column for further purification, from which fractions identified by SDS-PAGE were collected, pooled and concentrated to a final concentration of approximately 15 mg/ml. The concentrated wild type TcGPXI protein used to set up crystallisation screens at room temperature using various unbuffered ammonium sulphate concentrations (1.7 M - 2.1 M) in the absence of reducing agent and incubated at 20 °C. Well-defined crystals of the oxidised form of wild type TcGPXI, of prism morphology and approximately 0.2 mm were observed in conditions of unbuffered 1.8 M, 1.9 M and 2.0 M ammonium sulphate and were used for diffraction by an X-ray beam. X-ray diffraction data was collected to at a resolution of 2.3 Å and molecular replacement was carried out using the homology model of wild type TcGPXI (Chapter 5).

3.8 Summary

The DNA constructs for TcGPXI, TcGPXII, TbPxI and TbPxII were cloned into the pProEx vector. The constructs $TcGPXI_{14-177}$ (wild type), $TbPxI_{1-166}$, $TbPxII_{5-169}$, TcGPXIIa $_{69-225}$, TcGPXIIb $_{5-160}$ were selected for structural studies and the recombinant proteins expressed in *E. coli* Rosetta(DE3)pLysS bacterial host strain in order to optimise protein yield sufficient for structural studies. The purification of these proteins by immobilised metal affinity chromatography yielded soluble proteins that were subjected to additional purification stages using size exclusion chromatography and found to be monomeric. The proteins were then used to carry out crystallisation trials.

Sequence analyses of the primary sequences of all the proteins predict the presence of a glutathione peroxidase domain. The primary sequence of TcGPXII is shown to be more similar to human PHGPX4, whereas TcGPXI and TbPxI/TbPxII show closer homology to PtGPX5. These proteins adopt a thioredoxin fold (overall topology α-β-α sandwich), characteristic of the GPX family. The homology models for TbPxI and TcGPXI suggest the presence of similar catalytic redox centres and catalytic mechanisms with the presence of a disulphide bond in the oxidised form, resulting from the unwinding of the functional helix (α4) present in the reduced form. The TcGPXII homology model would suggest this enzyme to follow a mechanism characteristic of the classic GPX enzymes (use of a single catalytic cysteine), analogous to the PHGPX enzyme, however this remains ambiguous and would need to be determined experimentally.

Schlecker et al., 2007 have shown the formation of a disulphide bond in the oxidised form in the catalytic mechanism of the enzyme, supporting the possibility of a similar mechanism in TbPxI and TbPxII (these sequences show approximately 95% sequence identity to each other). Similarly, the homology models of wild type TcGPXI in the reduced form show the presence of a helix that is suggested to unwind and facilitate the formation of a disulphide bond in the oxidised form.

Site-directed mutagenesis was carried out to mutate cysteine residues of wild type TcGPXI to glycine (C48G, C77G, C96G) and the resultant proteins were expressed and purified to carry out NMR and kinetic investigation (discussed in Chapter 4). The purified C48G and C96G mutant proteins were also subjected to crystallisation studies. Crystalline material of wild type TcGPXI enzyme was

observed in conditions of unbuffered 2.0 M ammonium sulphate and additionally 2.0 M ammonium sulphate, 20 % (w/v) PEG-3000, 100 mM Tris pH 7.0 and 200 mM calcium acetate. TcGPXI C96G mutant enzyme crystals were also observed in conditions of 2.0 M ammonium sulphate, 0.2 M lithium sulphate and 0.1 M Tris pH 7.0. In-house crystallisation screens were designed and implemented using the hanging drop method in order to reproduce and optimise crystallisation, from which substrate treated wild type TcGPXI crystals in the oxidised form were obtained in conditions of 1.8 - 2.0 M ammonium sulphate and found to be suitable for X-ray diffraction (Chapter 5).

Chapter Four

Results

NMR SPECTROSCOPY AND ENZYMATIC ASSAYS DESCRIBING A PUTATIVE CATALYTIC MECHANSIM

4.1 Nuclear magnetic resonance spectroscopy (NMR)

The nuclear magnetic resonance phenomenon, first observed in 1946 has become a widespread tool for the non-destructive analysis of proteins. NMR techniques can provide information on structure, conformation and internal mobility. Comprehensive introductions include Sanders and Hunter (1987) and Derome (1987) (Sanders and Hunter, 1987; Derome, 1987).

The event of the NMR experiment occurs when nuclei are placed in a static magnetic field and then perturbed by a second, orthogonal oscillating magnetic field. Nuclei that possess the property of magnetic spin, quantum number 1/2, align with or against the direction of the external static magnetic field. The equilibrium bulk magnetization aligns with the external field, designated the lower energy state. When an oscillating radiofrequency is applied, the spins will undergo a transition to the higher energy state and the subsequent relaxation back to equilibrium results in RF radiation. The signal generated and that can be detected, known as the Free Induction Decay (FID), is a function of time, which upon Fourier transformation is converted into a domain of frequencies. Such emitted radiation can then be used to generate a one-dimensional spectrum (1D). The energy of the NMR transition (E) depends on the magnetic field strength (B_0) and a proportionality factor for each nucleus called the magnetogyric ratio (γ), *i.e.* $E = \gamma B_0$. The local environment around any given nucleus, such as the neighbouring nuclei, will affect the exact transition energy.

The frequency of NMR signals observed in a 1D spectrum is denoted by chemical shift values (parts per million or ppm). The chemical shift scale is referenced to a standard compound (such as 2,2-dimethylsilapentane-5-sulfonic acid (DSS) or tetramethylsilane (TMS)), whose methyl resonances are defined as 0 ppm, therefore resonances are defined by individual frequency differences from the reference peak (eliminating any magnetic field dependency).

The frequency of NMR signals observed depends on the nuclear isotope under investigation. Typically in studies of biomolecular macromolecules these isotopes include ^{1}H, ^{13}C, ^{15}N and ^{31}P. While ^{1}H has a high natural abundance (99.98 %), the isotopes ^{13}C and ^{15}N occur naturally at a lower abundance of 1.1% and 0.37% respectively, and are therefore incorporated into proteins to facilitate NMR experiments. A typical 1D ^{1}H spectrum will allow the observation of whether a protein contains globular folded components but further information is limited due to

complexity of overlapping resonances. The procedure of uniformly labelling proteins with the isotopes ^{13}C and ^{15}N allows two and three dimensional spectra to be recorded which correlates information between the different nuclei from the transfer of magnetization through either bonds, scalar coupling, or through space, dipolar coupling.

In this study only the application of NMR spectroscopy used with regard to scalar coupling to deduce structural information relevant to this protein is discussed, rather than the in-depth complex nature of the theory and mathematics behind it.

4.2 One dimensional ^{1}H NMR of wild type TcGPXI

1D ^{1}H NMR experiments allow the determination of whether a protein adopts a globular fold.

The frequencies at which the ^{1}H NMR resonances appear, chemical shift values, are dependent on the chemical group to which the proton is attached. For a protein that is unfolded every chemical group, *e.g.* a threonine methyl group will have a signal at a similar frequency. In the presence of a globular folded protein, resonances depend on both the bonded electrons and the non-bonded electrons of the atoms that are physically close in space. These neighbouring atoms create a unique electronic environment thus signals for threonine methyl groups are no longer typically equivalent but dispersed over a range of frequencies. The neighbouring electronic cloud can either "shield" or "deshield" a nucleus from the applied magnetic field (B_0). If the nuclei are shielded the resonance frequency is lower and the signal shifts to the right of the spectrum, whereas if there is deshielding the resonance frequency is higher and peaks appear to the left of the spectrum.

Typically, when a protein containing stable globular folded components is subjected to a 1D ^{1}H NMR experiment the resulting spectrum shows resonances dispersed over a broad range of frequencies (Figure 4.1a), whereas an unfolded protein exhibits peaks over a smaller range (Figure 4.1b).

An unfolded polypeptide such as the protein Hax1 (present in our laboratory), typically shows chemical shifts for methyl groups in the range of 0.8 - 1.2 ppm and for the amide protons 7.0 - 8.5 ppm (grey regions in Figure 4.1).

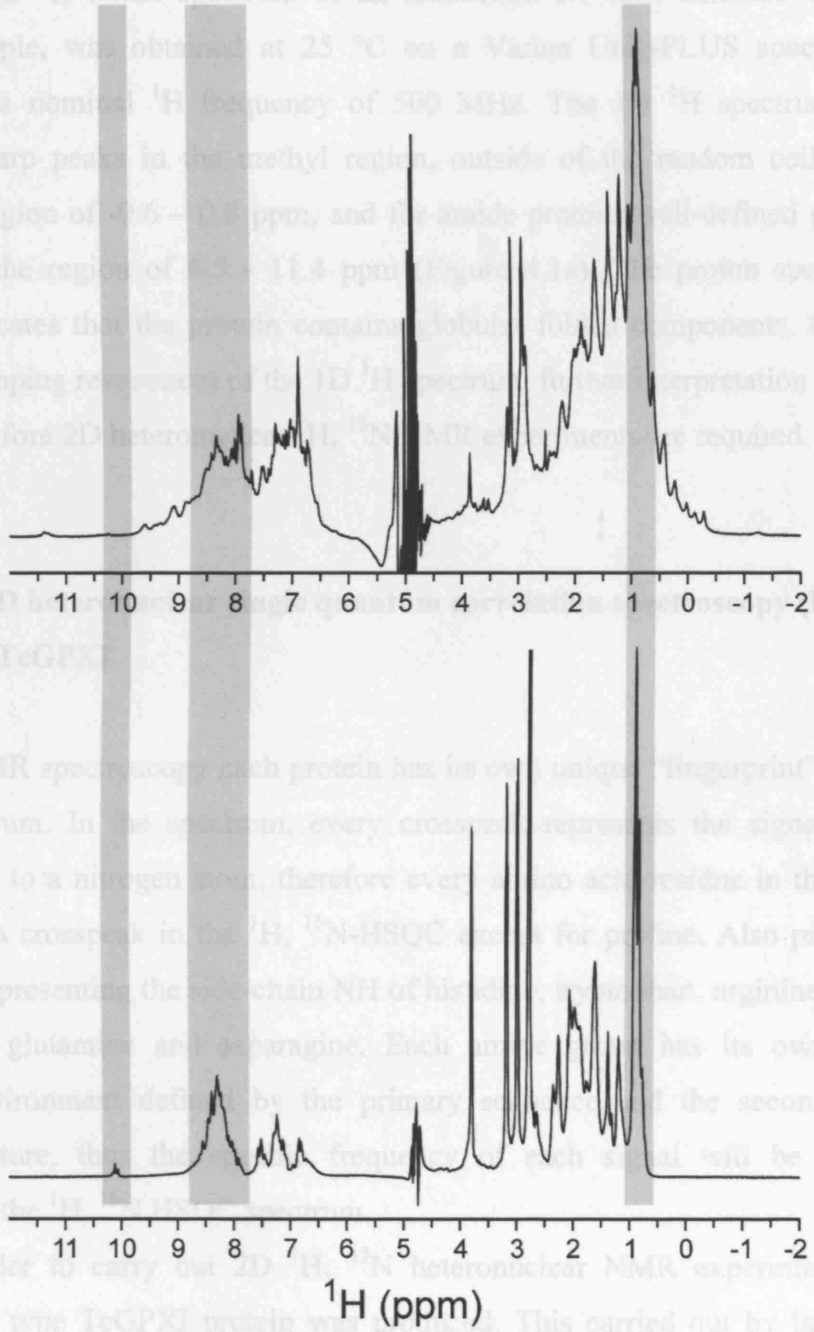


Figure 4.1: 1D ^{1}H NMR spectra

(a) wild type TcGPXI in 50mM sodium phosphate pH6.5 and 200mM NaCl.

(b) Hax1 (an unfolded protein).

The grey boxes illustrate random coil regions for methyl group resonances (~0.8 ppm), NH groups (7.5 - 8.5 ppm) and tryptophan side-chain NH signals (~10.2ppm).

The 1D ^{1}H NMR spectrum of an unlabelled 0.7 mM oxidised wild type TcGPXI sample, was obtained at 25 °C on a Varian UnityPLUS spectrometer, operating at a nominal ^{1}H frequency of 500 MHz. The 1D ^{1}H spectrum shows dispersed, sharp peaks in the methyl region, outside of the random coil section, spanning a region of -0.6 – 0.8 ppm, and for amide protons well-defined peaks are observed in the region of 6.5 - 11.4 ppm (Figure 4.1a). The proton spectrum of TcGPXI indicates that the protein contains globular folded components. However, due to overlapping resonances of the 1D ^{1}H spectrum, further interpretation of data is difficult therefore 2D heteronuclear ^{1}H, ^{15}N NMR experiments are required.

4.3 ^{1}H, ^{15}N 2D heteronuclear single quantum correlation spectroscopy (HSQC) of wild type TcGPXI

In NMR spectroscopy each protein has its own unique "fingerprint" ^{1}H, ^{15}N-HSQC spectrum. In the spectrum, every crosspeak represents the signal from a proton bound to a nitrogen atom, therefore every amino acid residue in the protein should have a crosspeak in the ^{1}H, ^{15}N-HSQC except for proline. Also present are crosspeaks representing the side-chain NH of histidine, tryptophan, arginine residues and NH_2 of glutamine and asparagine. Each amide group has its own unique chemical environment defined by the primary sequence and the secondary and tertiary structure, thus the specific frequency of each signal will be uniquely positioned in the ^{1}H, ^{15}N HSQC spectrum.

In order to carry out 2D ^{1}H, ^{15}N heteronuclear NMR experiments, ^{15}N-labelled wild type TcGPXI protein was produced. This carried out by large scale growth of *E. coli* and overexpression in minimal media containing ^{15}N-ammonium sulphate $(^{15}NH_4)_2SO_4$ as the sole nitrogen source, thus resulting in a protein that is isotopically enriched with ^{15}N. The protein was purified in the same manner as unlabelled protein (Chapter 2). Sufficient protein for 2D ^{1}H, ^{15}N NMR experiments was obtained from four litres of cell culture, grown and expressed in isotopically enriched minimal media yielded. A final oxidised wild type TcGPXI sample concentrated to 0.7 mM in a buffer of 50mM phosphate pH 6.5, 300mM NaCl and 10% D_2O. The ^{1}H, ^{15}N HSQC spectrum was acquired at 25 °C and is shown in Figure 4.2.

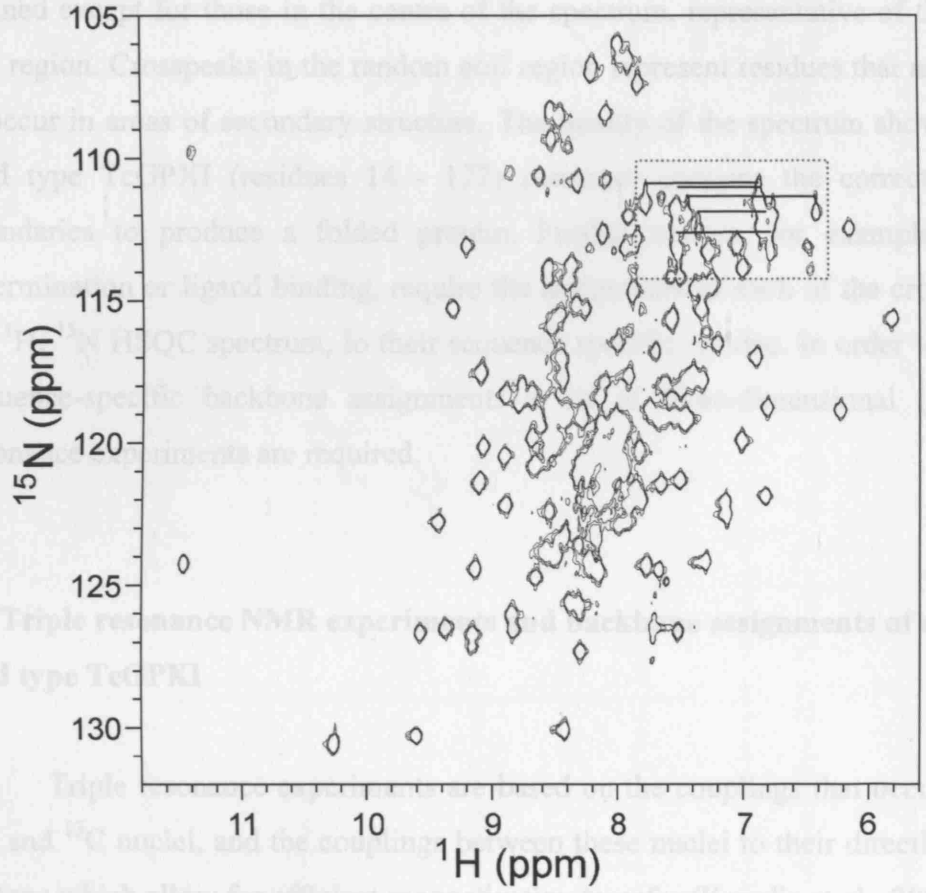


Figure 4.2: ^{1}H, ^{15}N HSQC (2D) spectrum of wild type TcGPXI

2D ^{1}H, ^{15}N HSQC spectrum of ^{15}N-labelled wild type TcGPXI in 50mM sodium phosphate, pH 6.5, 300mM NaCl.

The blue boxed region indicates the region containing signals from NH_2 side-chain of Asn, Gln residues.

The ^{1}H, ^{15}N HSQC spectrum shows crosspeaks that are well dispersed, of similar linewidth and of homogenous intensity. The 2D ^{1}H, ^{15}N-HSQC contains in total ca. 170 peaks, with a maximum of 154 peaks from backbone amide groups (*i.e.* total 164 residues- 10 prolines). The intensities of the crosspeaks are strong and defined except for those in the centre of the spectrum, representative of the random coil region. Crosspeaks in the random coil region represent residues that are unlikely to occur in areas of secondary structure. The quality of the spectrum shows that the wild type TcGPXI (residues 14 - 177) construct contains the correct sequence boundaries to produce a folded protein. Further studies, for example structure determination or ligand binding, require the assignment of each of the crosspeaks in the ^{1}H, ^{15}N HSQC spectrum, to their sequence specific residue. In order to carry out sequence-specific backbone assignments a set of three-dimensional (3D) triple resonance experiments are required.

4.4 Triple resonance NMR experiments and backbone assignments of oxidised wild type TcGPXI

Triple resonance experiments are based on the couplings that occur between ^{15}N and ^{13}C nuclei, and the couplings between these nuclei to their directly attached protons which allow for efficient magnetization transfer (Kanelis et al., 2001). Triple resonance experiments used in this study are HNCA, HN(CO)CA, HN(CA)CO, HNCO, HNCACB, CBCA(CO)NH, HA(CA)NH and HA(CACO)NH. A combination of some or all of the triple resonance experiments may be used to deduce ^{1}H and ^{15}N backbone amide assignments. The nomenclature of the triple resonance experiments describes the magnetization pathway for the specific experiment (Harris and Driscoll, 2007). For example, the HN(CA)CO experiment correlates the amide ^{1}H and ^{15}N resonances of one residue (i), with the CO chemical shifts of both the preceding residue (i-1) and the residue itself. The names of the nuclei written in parenthesis - the Cα group in this example - are only used for magnetization transfer and their specific frequencies are not detected. A diagrammatic representation of each of the triple resonance experiments and the magnetization transfers that occur is shown in Figure 4.3.

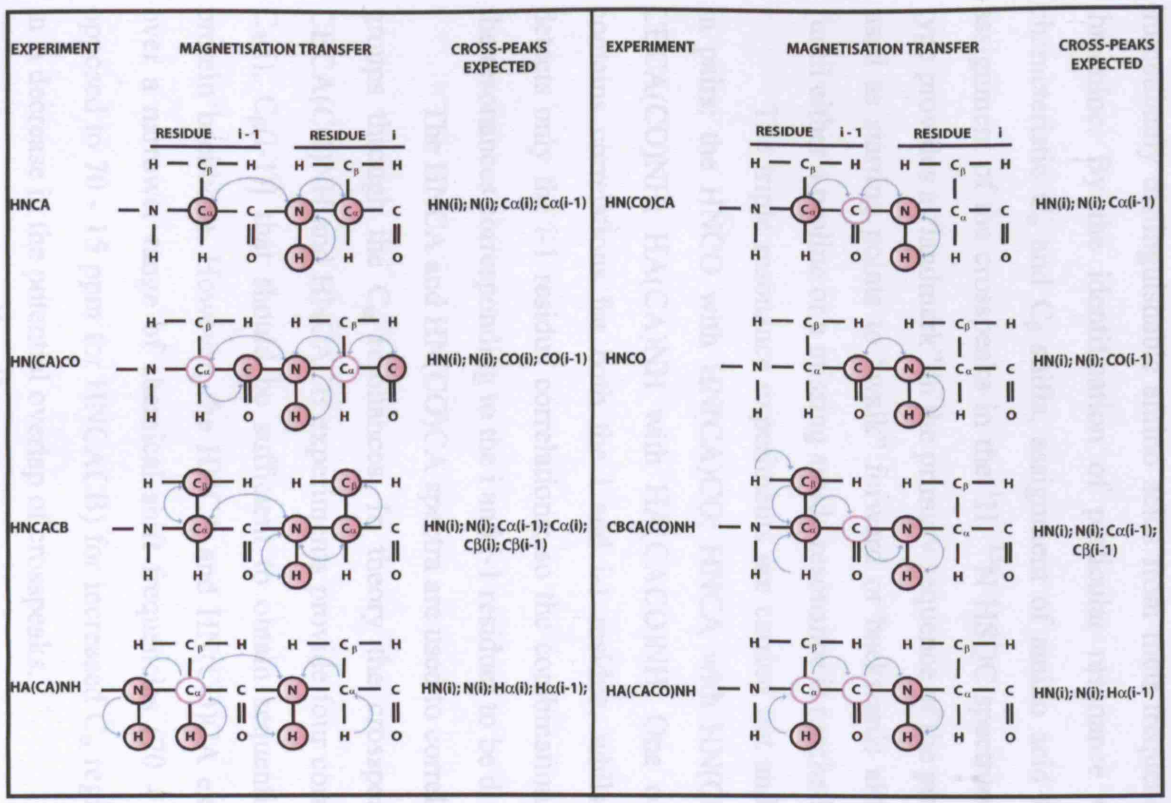


Figure 4.3: Experimental strategy for resonance assignment of backbone HN, N, Cα, Cβ, Hα and CO nuclei in a dipeptide fragment showing residues i and i-1

The path of magnetisation transfer in the triple resonance experiments HNCA, HN(CO)CA, HNCO, HN(CA)CO, CBCA(CO)NH, HNCACB, HA(CA)NH and HA(CACO)NH experiments are shown by arrows. The chemical shift of spins that are recorded are shown in pink filled circles and spins whose chemical shifts are not mapped are placed in parentheses in the pulse sequence name and outlined in a pink circle.

Using the data obtained from a complete set of triple resonance experiments, it is possible to correlate the H^N and N nuclei of residue i with its intra-residue (i) and preceding residue (i-1) C_α, H_α, CO and side chain C_β nuclei. A good starting point in backbone assignment is an amino acid with a characteristic combination of C_α and C_β chemical shifts. For example, glycine has C_α chemical shift ~45 ppm, no C_β resonance; alanine, which has frequencies of C_α ~53 ppm and C_β ~19 ppm. Other immediately distinguishable amino acids from their frequency range are serine and threonine. By the identification of particular resonance values of amides with characteristic C_α and C_β shifts, assignment of amino acid type can be made. The assignment of the crosspeaks in the ^1H, ^{15}N HSQC spectrum to a given amino acid type provides a "landmark" in the primary sequence of the protein, which can then be used as starting points to "walk" forward or backwards along the backbone chain (until either a proline or a missing amide resonance is reached).

The triple resonance experiments are carried out and subsequently analysed in pairs: the HNCO with HN(CA)CO; HNCA with HN(CO)CA; HNCACB with CBCA(CO)NH; HA(CA)NH with HA(CACO)NH. One of the two experiments contains correlations for both the i and i-1 residue, whilst the other experiment detects only the i-1 residue correlations; so the combination of both spectra allows the resonances corresponding to the i and i-1 residues to be distinguished.

The HNCA and HN(CO)CA spectra are used to correlate neighbouring amide groups through the C_α resonances. In theory the crosspeaks recorded from the CBCA(CO)NH and HNCACB experiments provide four correlations [C_α(i), C_α(i-1), C_β(i), C_β(i-1)] that should be sufficient to obtain sequential assignments for the protein backbone. However, the HNCA and HN(CO)CA experiments are acquired over a narrower range of chemical shift frequencies (70 - 45 ppm for HNCA as opposed to 70 - 15 ppm for HNCACB) for increased C_α region resolution, resulting in a decrease in the potential overlap of crosspeaks.

On a 0.8 mM [^{15}N, ^{13}C]-labelled sample of oxidised wild type TcGPXI, in a 50mM phosphate buffer pH 6.5, 10% D_2O and 300mM NaCl, the following set of triple resonance experiments were acquired: HNCA, HN(CO)CA, HNCACB, CBCA(CO)NH, HN(CA)CO, HNCO, HA(CA)NH, HN(CO)NH. The parameters used to acquire triple resonance data are listed in section (Chapter 2). The program ANSIG (Kraulis, 1989) was used to analyse the spectra and to correlate $^1H/^{15}N$

resonances to $^{13}C/^{1}H_{\alpha}$ resonances, in order to facilitate the sequential assignments of backbone residues. From the analysis of the spectra of a specific residue (i) crosspeaks from $C\alpha_i$, $C\alpha_{i-1}$, $C\beta_i$, $C\beta_{i-1}$, $H\alpha_i$, $H\alpha_{i-1}$ will be present. An example of how specific assignments were made using the triple resonance spectra for the amide groups Gln68 to Val72 is shown in Figure 4.4. For V72 there are four correlations shown in Figure 4.4 corresponding to the $C\alpha_{(i, i-1)}$ and $C\beta_{(i, i-1)}$ resonances. The $C\alpha_{i-1}$ and $C\beta_{i-1}$ can be identified by the correlations in the HN(CO)CA, coloured red, and the CBCA(CO)NH, in green. The $C\beta_{i-1}$ chemical shift (70.6 ppm) being at a higher ppm than the $C\alpha_{i-1}$, is characteristic of a threonine or, less likely, a serine. Looking through all the other amide correlations allows the identification of an amide that has matching Cα and Cβ chemical shifts, but as i-residue correlations so are observed only in the HNCA (black) and HNCACB (blue). If there is only one possible amide that has matching chemical shifts, then it may be assigned as T71, otherwise the Hα correlations (Figure 4.4.b) can be used to overcome any ambiguity as to which amide resonances is T71. The assignment can then be followed down the backbone as can be seen in this example for F70, G69 and Q68.

Complete analysis of the data recorded from triple resonance experiments [^{15}N, ^{13}C] oxidised wild type TcGPXI allowed the definitive assignments of 148 out of 154 (96% completed amide resonance assignment) amino acid residues (total 164 residues less 10 prolines). The assigned ^{1}H, ^{15}N-HSQC of oxidised wild type TcGPXI is shown in Figure 4.5.The backbone amides that could not be assigned in the 2D ^{1}H, ^{15}N HSQC spectrum were: Cys48, Asn78, Thr97, Gly113, Ser114 and K134. Despite being unable to assign the position of the amide groups of these residues in the 2D ^{1}H, ^{15}N HSQC spectrum, a total of 99% of $^{13}C_{\alpha}$ chemical shifts, 91% of $^{13}C_{\beta}$ chemical shifts, 97% of ^{13}CO chemical shifts and 99% of $^{1}H_{\alpha}$ chemical shifts were assigned. The stable oxidised wild type shows unusual amide chemical shifts for residues T71, F103, L119 and W138 whose crosspeaks are highlighted in Figure 4.5.

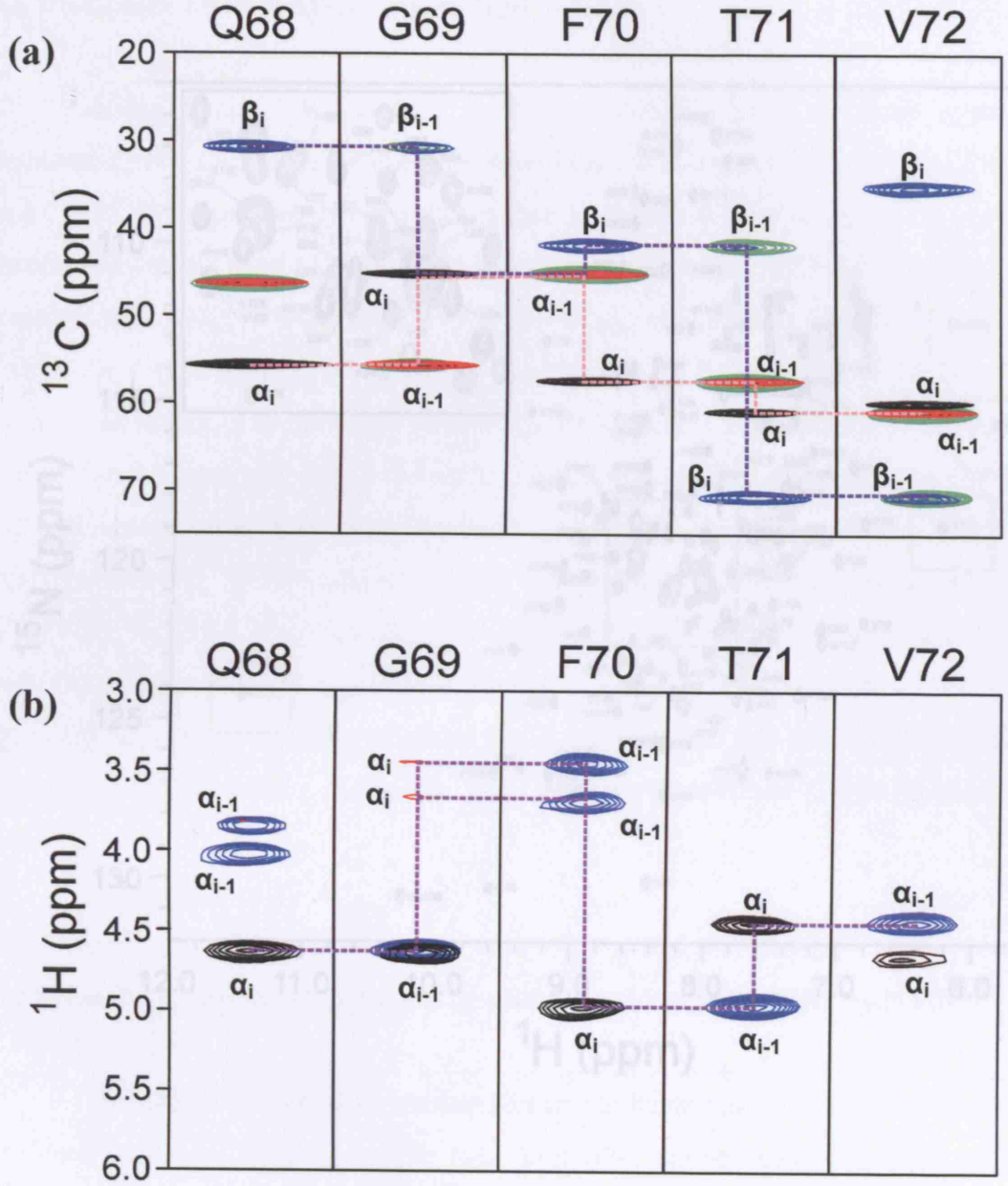


Figure 4.4: Strip plots showing correlations from residues Q68 to V72

(a) HNCA (black), HNCACB(blue), HN(CO)CA(red), CBCACONH(green) spectra showing examples of $^1H^N/^{15}N \rightarrow {}^{13}C$ correlations. Cross peaks are labelled as originating from α or β, intra- (i) and inter (i-1) carbon atom correlations.

(b) HA(CA)NH (black: positive contours; red: negative contours) and HA(CO)NH (blue) showing $^1H^N/^{15}N \rightarrow {}^1H_\alpha$ correlations.

The dotted lines illustrate sequential assignment of selected $^1H^N/^{15}N \rightarrow {}^{13}C\alpha_i$, $^{13}C\alpha_{i-1}$,(pink); $^{13}C\ \beta_i$, $^{13}C\beta_{i-1}$, (purple); $^1H\alpha_i$, and $^1H\alpha_{i-1}$ correlations (purple).

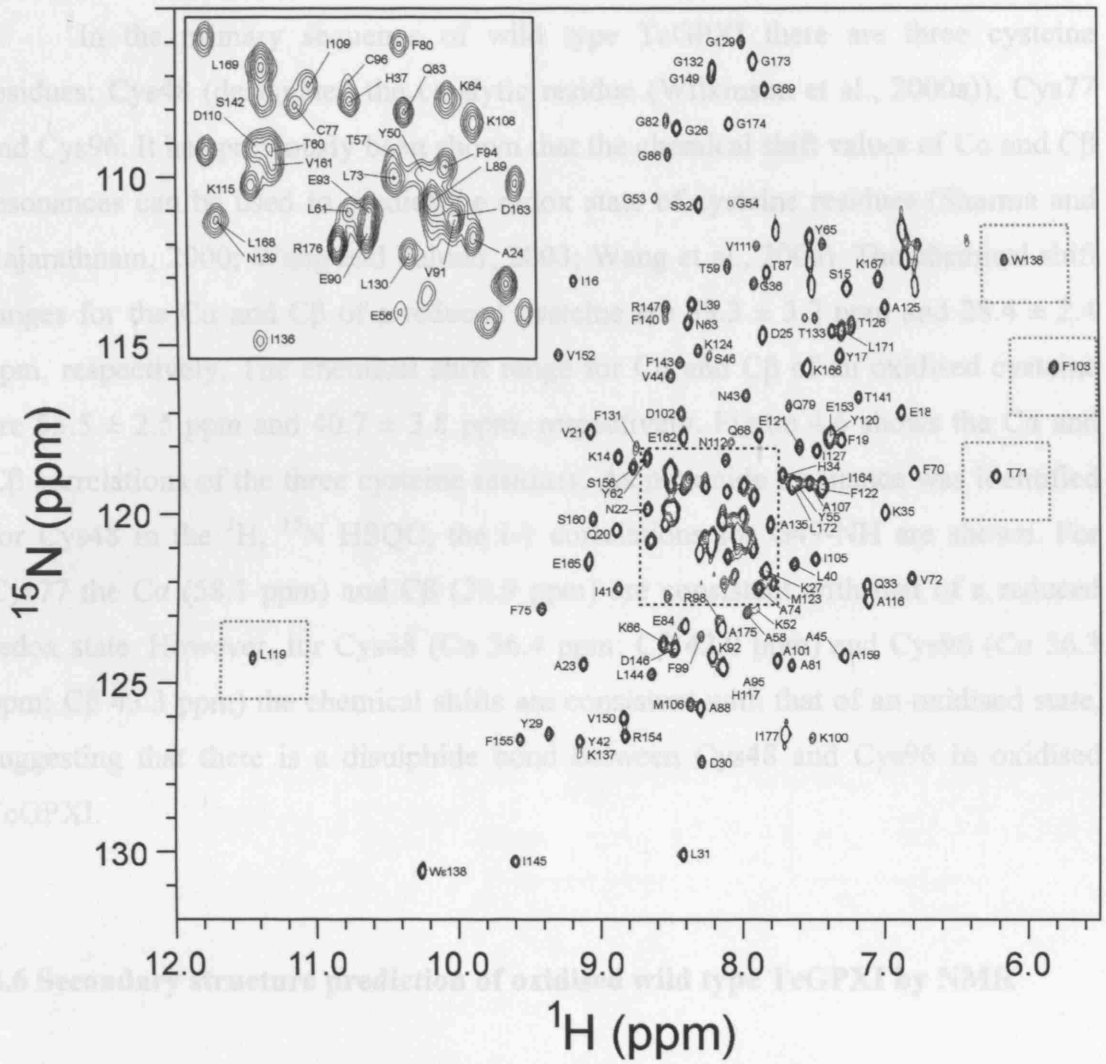


Figure 4.5: NMR sequence specific assignments of oxidised wild type TcGPXI

^{1}H, ^{15}N HSQC spectrum of ^{13}C, ^{15}N-labelled wild type TcGPXI in 50mM sodium phosphate, pH 6.5, 300 mM NaCl buffer.

Sequence-specific backbone assignments for all residues except Cys48, Asn78, Thr97, Gly113, Ser114 and K134. Side-chain NH group of tryptophan is denoted Wε138.Residues showing unusual chemical shifts are shown in red boxes (Leu119, Phe103, Thr71 and Trp138).

4.5 Disulphide bond in oxidised wild type TcGPXI

In the primary sequence of wild type TcGPXI there are three cysteine residues: Cys48 (designated the catalytic residue (Wilkinson et al., 2000a)), Cys77 and Cys96. It has previously been shown that the chemical shift values of Cα and Cβ resonances can be used to predict the redox state of cysteine residues (Sharma and Rajarathnam, 2000; Wang and Palmer, 2003; Wang et al., 2006). The chemical shift ranges for the Cα and Cβ of a reduced cysteine are 59.3 ± 3.2 ppm and 28.4 ± 2.4 ppm, respectively. The chemical shift range for Cα and Cβ of an oxidised cysteine are 55.5 ± 2.5 ppm and 40.7 ± 3.8 ppm, respectively. Figure 4.6 shows the Cα and Cβ correlations of the three cysteine residues. As no amide resonance was identified for Cys48 in the ^{1}H, ^{15}N HSQC, the i-1 correlations for G49-NH are shown. For Cys77 the Cα (58.1 ppm) and Cβ (29.9 ppm) are consistent with that of a reduced redox state. However, for Cys48 (Cα 56.4 ppm; Cβ 42.0 ppm) and Cys96 (Cα 56.3 ppm; Cβ 43.3 ppm) the chemical shifts are consistent with that of an oxidised state, suggesting that there is a disulphide bond between Cys48 and Cys96 in oxidised TcGPXI.

4.6 Secondary structure prediction of oxidised wild type TcGPXI by NMR

The chemical shift data obtained from the triple resonance experiments on oxidised wild type TcGPXI can be used to predict the secondary structure of the protein. It has long been established that the H_α, C_α, C_β and CO chemical shifts in secondary structural elements relative to chemical shifts in random coil, have a clear correlation with the polypeptide torsion backbone angles φ and ψ (Spera and Bax, 1991; Wishart and Sykes, 1994; Wishart et al., 1991). The chemical shifts of C_α and carbonyl carbon nuclei experience an upfield shift for amino acids that occur in β-sheets and downfield shift for amino acids that occur in α-helices. The chemicals shift of $^{13}C_\beta$ and $^{1}H_\alpha$ nuclei show an opposite correlation *i.e.* there is an upfield shift from the random coil position when amino acids are present in a α-helical structure and a downfield shift when amino acids occur in a β-sheet. The individual C_α, C_β, $^{1}H_\alpha$ and CO chemical shifts observed were subtracted from the random coil shift

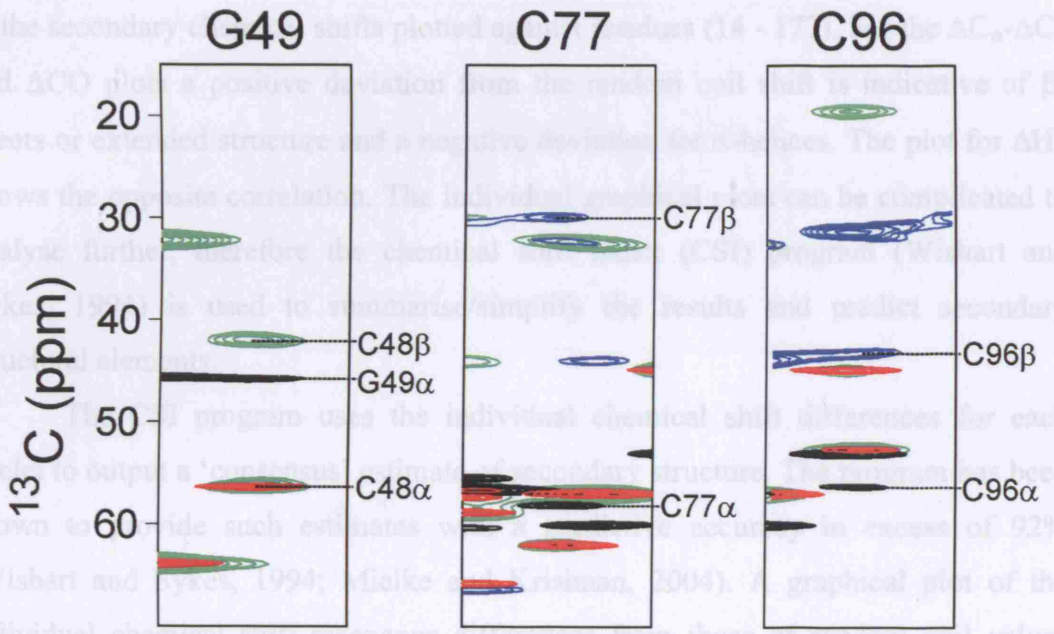


Figure 4.6: Strip plots for residues Gly49, Cys77 and Cys96 in oxidised wild type TcGPXI

Strip plots illustrating $^{13}C\alpha_i$ and $^{13}C\beta_i$ cross peaks for Cys77 and Cys96; and $^{13}C\alpha_{i-1}$, $^{13}C\beta_{i-1}$ for Gly49.

values *i.e.* $\Delta\delta = \delta$ r.c. (random coil chemical shift) – δ obs (observed chemical shift), to give the secondary chemical shifts ΔC_α, ΔC_β, ΔCO, ΔH_α the values of which can be plotted graphically. Since the deviations, ΔC_α and ΔC_β, are of similar magnitude and of opposite sign when in α-helix or β-sheet, the difference between the two secondary chemical shifts (ΔC_α-ΔC_β) enhances the correlation between the definition of the secondary structure and the chemical shift values. Figure 4.7a-c shows a plot of the secondary chemical shifts plotted against residues (14 - 177). For the ΔC_α-ΔC_β and ΔCO plots a positive deviation from the random coil shift is indicative of β-sheets or extended structure and a negative deviation for α-helices. The plot for ΔH_α shows the opposite correlation. The individual graphical plots can be complicated to analyse further, therefore the chemical shift index (CSI) program (Wishart and Sykes, 1994) is used to summarise/simplify the results and predict secondary structural elements.

The CSI program uses the individual chemical shift differences for each nuclei to output a 'consensus' estimate of secondary structure. The program has been shown to provide such estimates with a predictive accuracy in excess of 92% (Wishart and Sykes, 1994; Mielke and Krishnan, 2004). A graphical plot of the individual chemical shift resonance differences from those of random coil values shows positive deviation indicative of the presence of β-sheets and negative deviation indicative of the presence of α-helical structures. The software package Chemical Shift Index (CSI) (Wishart and Sykes, 1994) functions by comparing the individual observed chemical shifts of the H_α, C_α, C_β and CO for each residue to random coil conformation values. If the difference is positive for C_α and CO, and negative for C_β and H_α then the residue is assigned a +1 value; if the difference is negative for C_α and CO but positive for C_β and H_α nuclei then it is assigned a value of -1. Where there is no difference to the random coil conformation, a value of 0 is designated. The consensus secondary structure is deduced by interpreting these values of +1, -1 and 0. A dense grouping of greater than four +1 scores (uninterrupted) is designated an area of α-helical structure. Likewise, a dense grouping of greater than three -1 scores (uninterrupted) is designated as a β-strand. The resulting output is a graphical representation whereby residues are assigned values of +1, -1 and 0 indicative of β-strand, α-helix and random coil regions, respectively.

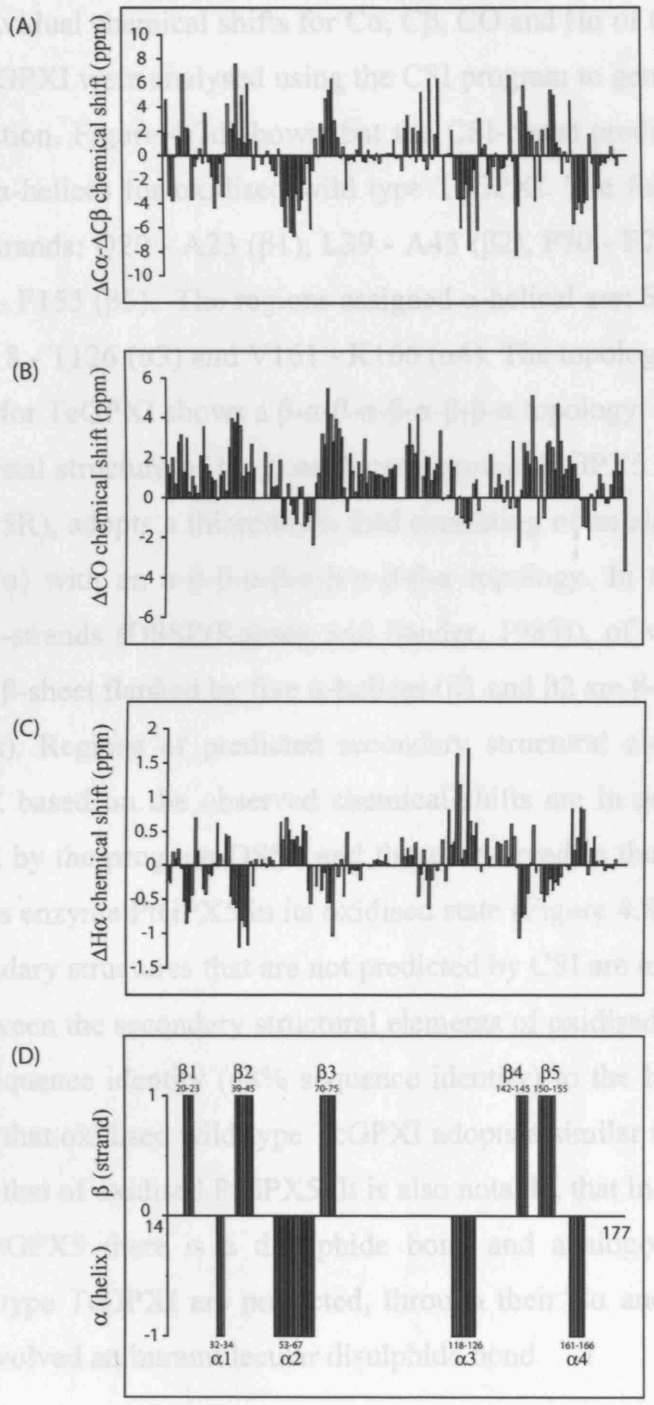


Figure 4.7: The chemical shift index for the secondary structure prediction of oxidised wild type TcGPXI

Secondary chemical shift values for: (a) ΔCα-ΔCβ, (b) ΔCO and (c) ΔHα.

(d) CSI (Wishart et al., 1992, Wishart and Sykes, 1994) secondary structure prediction for each residue based on Cα, Cβ, CO and Hα. β-strand regions are Q20 - A23 (β1); L39 - A45 (β2); F70 - F75 (β3); S142 - I145 (β4); V150 - F155 (β5). Regions assigned α-helical are S32 - H34 (α1); G53 - G67 (α2); P118 - T126 (α3); V161 - K166 (α4).

Blanks indicate absence of crosspeak or a residue that was not assigned.

The individual chemical shifts for Cα, Cβ, CO and Hα of the oxidised sample of wild type TcGPXI were analysed using the CSI program to generate the secondary structure prediction. Figure 4.7d shows that the CSI-based prediction contains 5 β-strands, and 4 α-helices for oxidised wild type TcGPXI. The following regions are assigned as β-strands: Q20 - A23 (β1), L39 - A45 (β2), F70 - F75 (β3), S142 - I145 (β4) and V150 - F155 (β5). The regions assigned α-helical are: S32 - H34 (α1), G53 - G67 (α2), P118 - T126 (α3) and V161 - K166 (α4). The topology obtained from the CSI prediction for TcGPXI shows a β-α-β-α-β-α-β-β-α topology.

The crystal structure of the homologous protein PtGPX5 in its oxidised form (PDB code: 2P5R), adopts a thioredoxin fold consisting of an alpha/beta/alpha layer sandwich (α/β/α) with an α-β-β-α-β-α-β-α-β-β-α topology. In the crystal structure there are six β-strands (DSSP(Kabsch and Sander, 1983)), of which β3-β6 form a central twisted β-sheet flanked by five α-helices (β1 and β2 are β-hairpin structures at the N-terminus). Regions of predicted secondary structural elements for oxidised wild type GPX based on the observed chemical shifts are in good agreement with those predicted by the program DSSP and those observed in the crystal structure of the homologous enzyme PtGPX5 in its oxidised state (Figure 4.8) (Koh et al., 2007). The only secondary structures that are not predicted by CSI are α1 and β2. The close agreement between the secondary structural elements of oxidised wild type TcGPXI, and its high sequence identity (48% sequence identity) to the homologue PtGPX5, would suggest that oxidised wild type TcGPXI adopts a similar thioredoxin fold and architecture to that of oxidised PtGPX5. It is also notable, that in the crystal structure of oxidised PtGPX5 there is a disulphide bond and analogous cysteines in the oxidised wild type TcGPXI are predicted, through their Cα and Cβ resonances, to similarly be involved an intramolecular disulphide bond.

4.7 Triple resonance experiments of reduced wild type TcGPXI

The crystal structures of TcGPXI homologue PtGPX5 reveal that it adopts different conformations in the oxidised (PDB code: 2P5R) and reduced forms (PDB code: 2P5Q). The crystal structure of the reduced form is shown to contain a helix which is suggested to unravel, and is a loop region in the oxidised conformation.

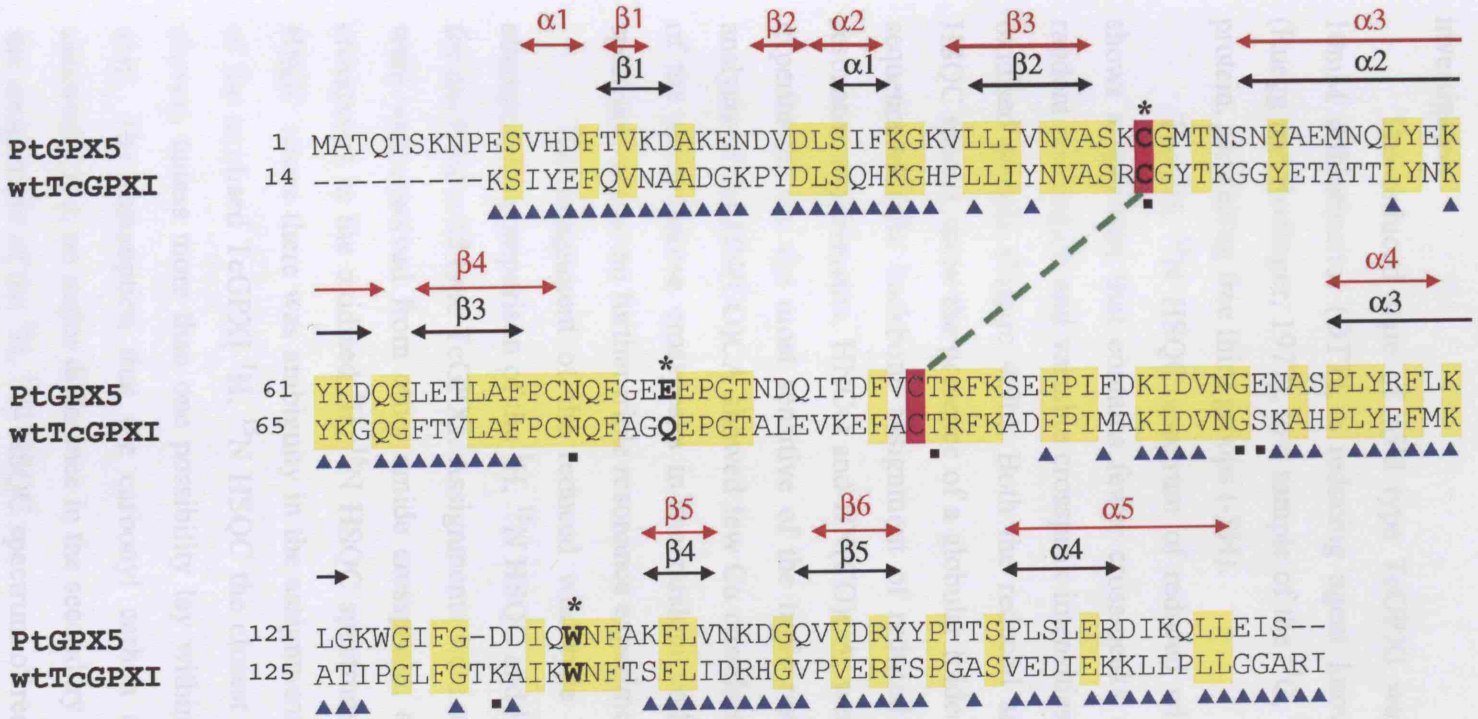


Figure 4.8: Sequence alignment of wild type TcGPXI and homologue PtGPX5

Secondary structure of oxidised PtGPX5 (Accession A3FNZ8), PDB: 2P5R (red). Secondary structure prediction (CSI) for wild type TcGPXI (black). Conserved catalytic residues of the GPX family (Cys, Gln/Glu, Trp) are denoted by (*). The dashed green line indicates a disulphide bond between cysteine residues (highlighted pink) present in oxidised PtGPX5 and predicted in oxidised wild type TcGPXI. Residues that were not assigned in the ^{1}H, ^{15}N HSQC of oxidised wild type TcGPXI are marked (■). Residues assigned in the ^{1}H, ^{15}N HSQC of reduced wild type TcGPXI are indicated (▲).

Published data shows that the reduced form of full length TcGPXI is achieved by the redox shuttle protein tryparedoxin, which in turn itself becomes oxidised. The reduced form of TcGPXI is then able to enzymatically reduce the hydroperoxide substrate. Therefore a reduced active state of wild type TcGPXI was investigated.

The reduced state of wild type TcGPXI was obtained by the addition of 10mM dithiothreitol (DTT), a reducing agent known to break disulphide bonds (Ruegg and Rudinger, 1977), to a sample of the ^{13}C, ^{15}N-labelled oxidised wild type protein, generating free thiol groups (-SH).

The ^{1}H, ^{15}N HSQC spectrum of reduced wild type TcGPXI (Figure 4.9a) shows a spectrum that contains fewer cross-peaks with broader linewidths in the random coil region and variable crosspeak intensities, when compared to that of the oxidised sample (Figure 4.9b). Both the reduced and oxidised wild type ^{1}H, ^{15}N HSQC spectra show the presence of a globular folded protein. In order to obtain the sequence-specific backbone assignment of reduced wild type TcGPXI two triple resonance experiments, HNCO and HN(CO)CA, were acquired initially. These two experiments are the most sensitive of the triple resonance experiments. However, analysis of the HN(CO)CA showed few Cα correlations in the 3D spectrum outside of the more intense crosspeaks in the random coil region of the ^{1}H, ^{15}N HSQC spectrum; and so no further triple resonance experiments were acquired.

The assignment of the reduced wild type TcGPXI protein was therefore attempted by comparison of the ^{1}H, ^{15}N HSQC and HNCO spectra to those obtained for oxidised wild type TcGPXI. Assignments were made for the NH crosspeaks that were well resolved from other amide crosspeaks, and that overlaid with assigned crosspeaks in the oxidised ^{1}H, ^{15}N HSQC spectrum. For crosspeaks in the ^{1}H, ^{15}N HSQC where there was ambiguity in the assignment, through the direct comparison of the oxidised TcGPXI ^{1}H, ^{15}N HSQC the closest match of carbonyl carbons was chosen, unless more than one possibility lay within 1 ppm of the carbonyl carbon shift. The assumption that the carbonyl carbon chemical shifts do not change drastically (*i.e.* no major difference in the secondary structure) may not hold true and the assignment of the ^{1}H, ^{15}N HSQC spectrum of reduced wild type TcGPXI is only tentative.

The sequence-specific backbone assignment of the reduced wild type

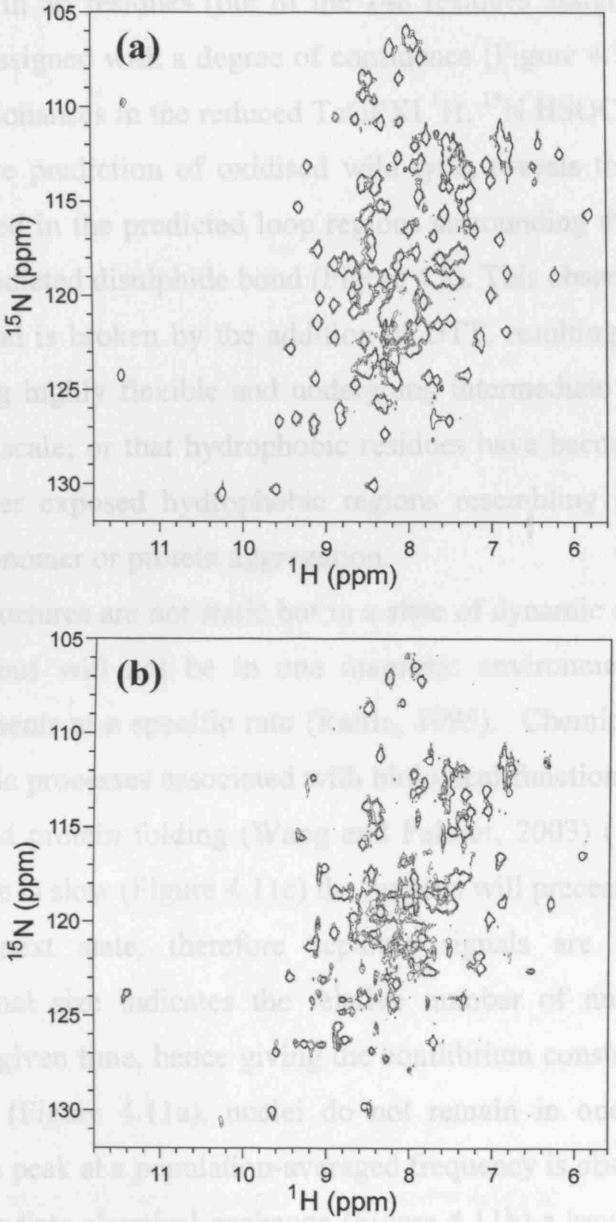


Figure 4.9: Comparison of ^{1}H, ^{15}N HSQC spectra of wild type TcGPXI

^{1}H, ^{15}N HSQC spectra of ^{15}N, ^{13}C-labelled wild type TcGPXI in the oxidised (a) and reduced (b) forms, in buffer 50mM sodium phosphate, pH 6.5, 300mM NaCl. The reduced form (b) additionally contains 10mM dithiotreitol (DTT).

spectrum resulted in 86 residues (out of the 148 residues assigned in the oxidised wild type) being assigned with a degree of confidence (Figure 4.10). Mapping those assigned amide resonances in the reduced TcGPXI ^{1}H, ^{15}N HSQC spectrum upon the secondary structure prediction of oxidised wild type, reveals that the residues not assigned are located in the predicted loop regions surrounding the cysteine residues involved in the predicted disulphide bond (Figure 4.8). This observation suggests that the disulphide bond is broken by the addition of DTT, resulting in the surrounding loop regions being highly flexible and undergoing intermediate chemical exchange on the NMR timescale; or that hydrophobic residues have become exposed and are packing with other exposed hydrophobic regions resembling a molecule slightly larger than the monomer or protein aggregation.

Protein structures are not static but in a state of dynamic equilibrium. In such instances, a nucleus will not be in one magnetic environment, but exchanging between environments at a specific rate (Rattle, 1995). Chemical exchange effects can reveal dynamic processes associated with biological functions including those of ligand binding and protein folding (Wang and Palmer, 2003) (Figure 4.11). When chemical exchange is slow (Figure 4.11c) the nucleus will precess many times before moving to the next state, therefore separate signals are observed for each environment (signal size indicates the relative number of nuclei in a particular environment and given time, hence giving the equilibrium constant). If the chemical exchange is fast (Figure 4.11a), nuclei do not remain in one environment long enough and thus a peak at a population-averaged frequency is observed.

In intermediate chemical exchange (Figure 4.11b) a broad coalesced peak is observed, which can broaden to such an extent that the signal is indistinguishable from noise, resulting in missing peaks in a 2D ^{1}H, ^{15}N HSQC spectrum. If a 2D ^{1}H, ^{15}N HSQC spectrum showed either more crosspeaks or a significantly reduced number of crosspeaks than expected, it would suggest that the protein sample is undergoing chemical exchange or conformational changes on the slow and intermediate timescale, respectively. Chemical exchange may also occur between the nuclei of exposed amino acid residues and the solvent, which can lead to reduced intensities of the crosspeaks or complete absence from the spectrum.

As an alternative to the addition of reducing agent to the oxidised protein to produce reduced TcGPXI, site directed mutagenesis of all three cysteine residues

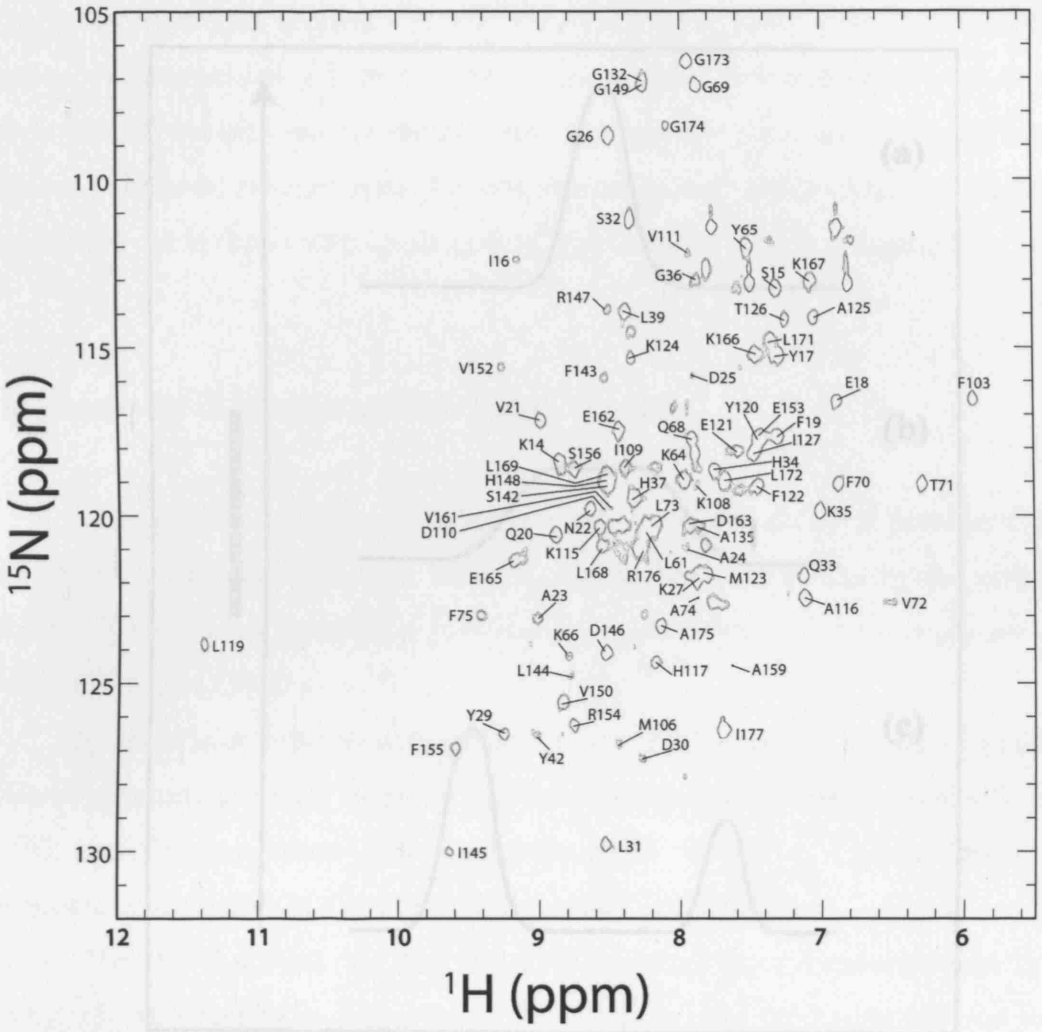


Figure 4.10: NMR sequential assignments of reduced wild type TcGPXI

^{1}H, ^{15}N HSQC spectrum of [^{13}C, ^{15}N]-labelled wild type TcGPXI in 50mM sodium phosphate, pH6.5, 300mM NaCl buffer. A total of 87 sequence specific amide resonances assigned.

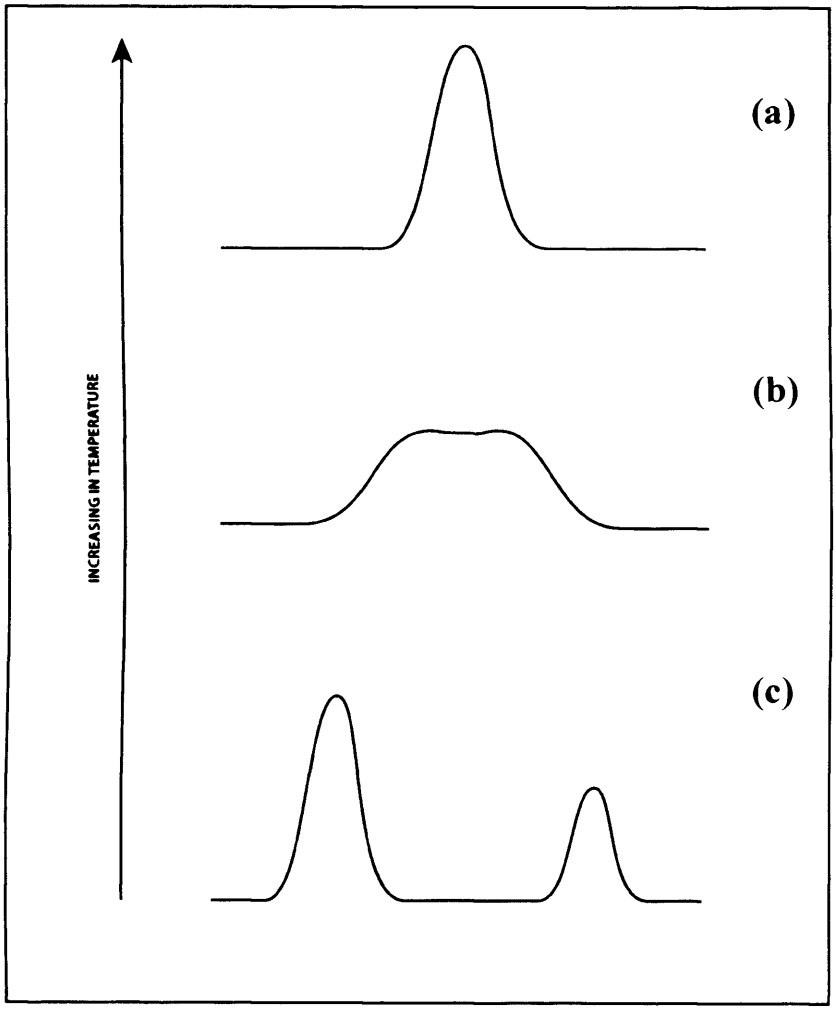


Figure 4.11: Diagrammatic representation of chemical exchange between two environments

(a) Fast exchange: nuclei do not remain in one environment long enough, thus an averaged frequency is recorded at the population-average position.

(b) Intermediate exchange: variation between the two extremes gives rise to a broad coalesced peak.

(c) Slow exchange: two peaks are observed representative of nuclei precessing many times in one environment before moving to the next environment.

present in the primary sequence of wild type TcGPXI (Cys48, Cys77 and Cys96) to glycine was carried out (Chapter 2). Mutations of the cysteine residues were carried out to identify which cysteine residues are important for the disulphide bond and to create a permanent reduced state. Protein expression and purification of the mutants was carried out in the same manner as wild type TcGPXI protein (Chapter 2).

4.8 Assessment of cysteine mutants by 1D ^{1}H NMR

The presence of a predicted intramolecular disulphide bond between Cys48 and Cys96 of oxidised wild type TcGPXI, was investigated further by the mutation of each of the three cysteine residues present in wild type TcGPXI to glycine, *i.e.* C48G, C77G and C96G.

In order to investigate whether the mutant protein samples contained globular folded components, the 1D ^{1}H NMR spectra of unlabelled 0.3mM samples of C48G, C77G and C96G mutants were acquired at 25 °C on a Varian UnityPLUS spectrometer, operating at a nominal ^{1}H frequency of 500 MHz.

The 1D ^{1}H spectra of C48G and C96G mutants show dispersed peaks in the methyl region spanning a region of -1.3 - 0.8 ppm, and for amide protons well-defined peaks are observed in the region of 6.5 - 11.4 ppm (Figure 4.12 a and c). The spectra indicate the presence of proteins containing globular folded components and are comparable to the 1D ^{1}H spectrum obtained from the wild type sample (Figure 4.1a). The 1D ^{1}H spectrum of the C77G mutant shows a broader spectrum in the amide proton region and less defined peaks, including those in methyl regions and amide proton region (Figure 4.12 b). In particular, no peak is observed at 11.3 ppm as compared to the 1D ^{1}H spectra of C48G and C96G mutants. The 1D ^{1}H spectrum of the C77G mutant, suggests that whilst comprising the characteristics of a globular folded protein, it contains broader linewidths that are suggestive of protein aggregation and so this mutant was not carried forward for further investigation.

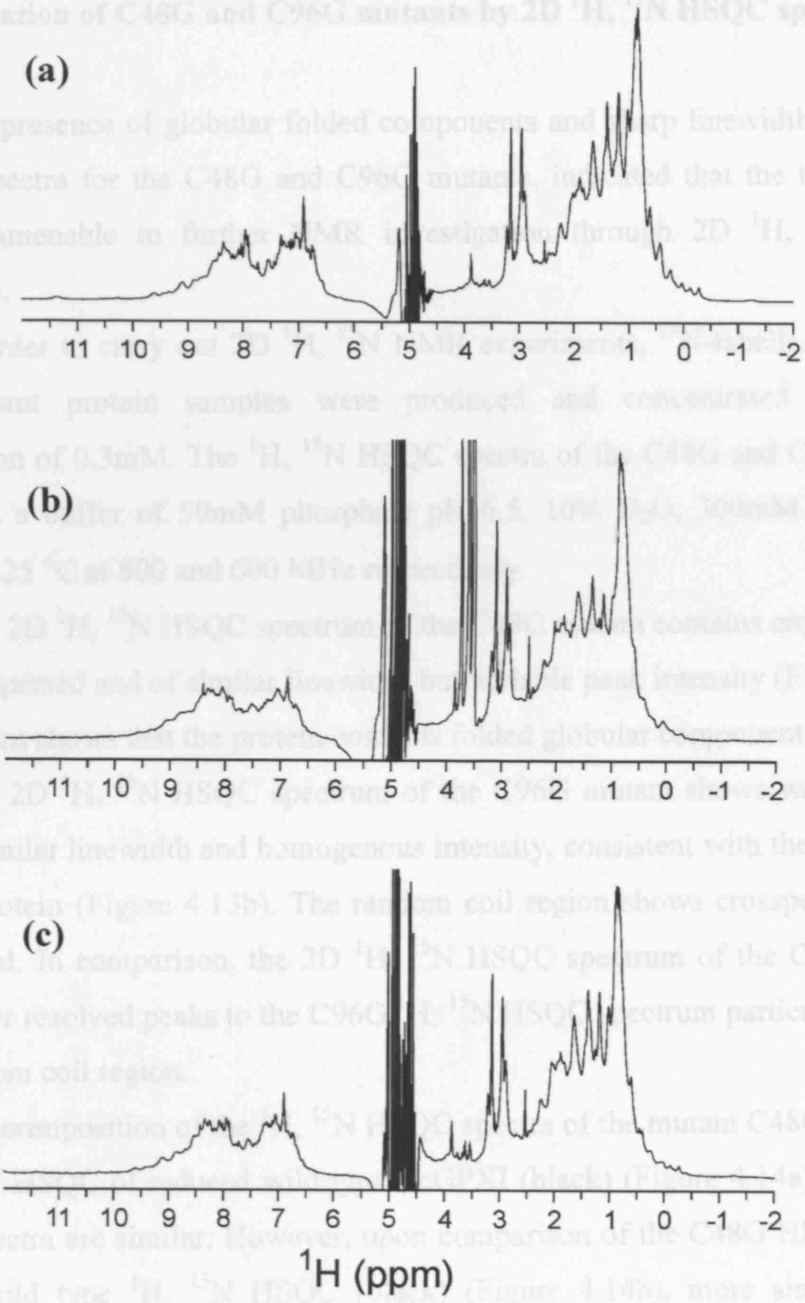


Figure 4.12: 1D ^{1}H spectra cysteine mutants of TcGPXI

1D ^{1}H spectra of (a) C48G, (b) C77G and (c) C96G mutant proteins, contained in buffer 50 mM sodium phosphate, pH 6.5, 200 mM NaCl. Spectra were obtained on a Varian UnityPLUS spectrometer operating at nominal ^{1}H frequency of 500 MHz at 25 °C.

4.9 Investigation of C48G and C96G mutants by 2D ^{1}H, ^{15}N HSQC spectroscopy

The presence of globular folded components and sharp linewidths in the 1D ^{1}H NMR spectra for the C48G and C96G mutants, indicated that the two mutants would be amenable to further NMR investigation through 2D ^{1}H, ^{15}N HSQC experiments.

In order to carry out 2D ^{1}H, ^{15}N NMR experiments, ^{15}N-labelled C48G and C96G mutant protein samples were produced and concentrated to a final concentration of 0.3mM. The ^{1}H, ^{15}N HSQC spectra of the C48G and C96G mutant samples, in a buffer of 50mM phosphate pH 6.5, 10% D_2O, 300mM NaCl, were acquired at 25 °C at 800 and 600 MHz respectively.

The 2D ^{1}H, ^{15}N HSQC spectrum of the C48G mutant contains crosspeaks that are well dispersed and of similar linewidth but variable peak intensity (Figure 4.13a). The spectrum shows that the protein contains folded globular components.

The 2D ^{1}H, ^{15}N HSQC spectrum of the C96G mutant shows well dispersed peaks of similar linewidth and homogenous intensity, consistent with the presence of a folded protein (Figure 4.13b). The random coil region shows crosspeaks that are well defined. In comparison, the 2D ^{1}H, ^{15}N HSQC spectrum of the C48G mutant shows fewer resolved peaks to the C96G ^{1}H, ^{15}N HSQC spectrum particularly shown in the random coil region.

Superimposition of the ^{1}H, ^{15}N HSQC spectra of the mutant C48G (red), with 2D ^{1}H, ^{15}N HSQC of reduced wild type TcGPXI (black) (Figure 4.14a), shows that the two spectra are similar. However, upon comparison of the C48G HSQC (red) to oxidised wild type ^{1}H, ^{15}N HSQC (black) (Figure 4.14b), more similarities are observed. It is clearly observed that the majority of peaks in the C48G mutant spectrum match with those of the oxidised wild type ^{1}H, ^{15}N HSQC spectrum, albeit with slight peak shift variation. This is specifically highlighted in the boxed regions of Figure 4.14, where there are more similar chemical shifts between the C48G mutant spectrum and the oxidised spectrum, than to the reduced spectrum. The C48G mutant spectrum is less well resolved in the random coil region so it is difficult to ascertain the exact number of conserved peaks in this region. Peaks that are exactly matching in the C48G mutant and oxidised wild type spectra, or exhibiting slight peak shifts or peak weakening in the mutant spectra, are more apparent in the

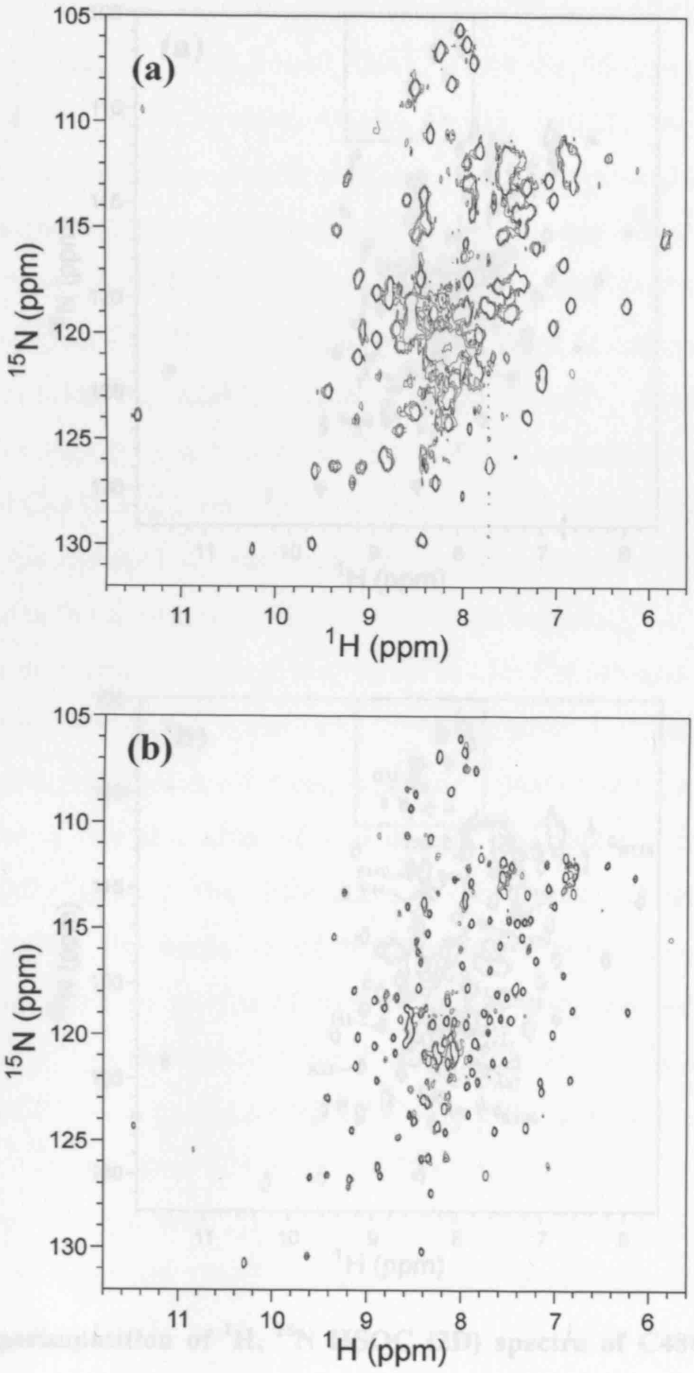


Figure 4.13: (2D) ^{1}H , ^{15}N HSQC spectrum of C48G and C96G TcGPXI mutants

2D ^{1}H, ^{15}N HSQC spectrum of ^{15}N-labelled C48G (a) and C96G (b) TcGPXI mutants in 50 mM sodium phosphate, pH 6.5, 300 mM NaCl acquired at 600 MHz (C48G) and 800 MHz (C96G).

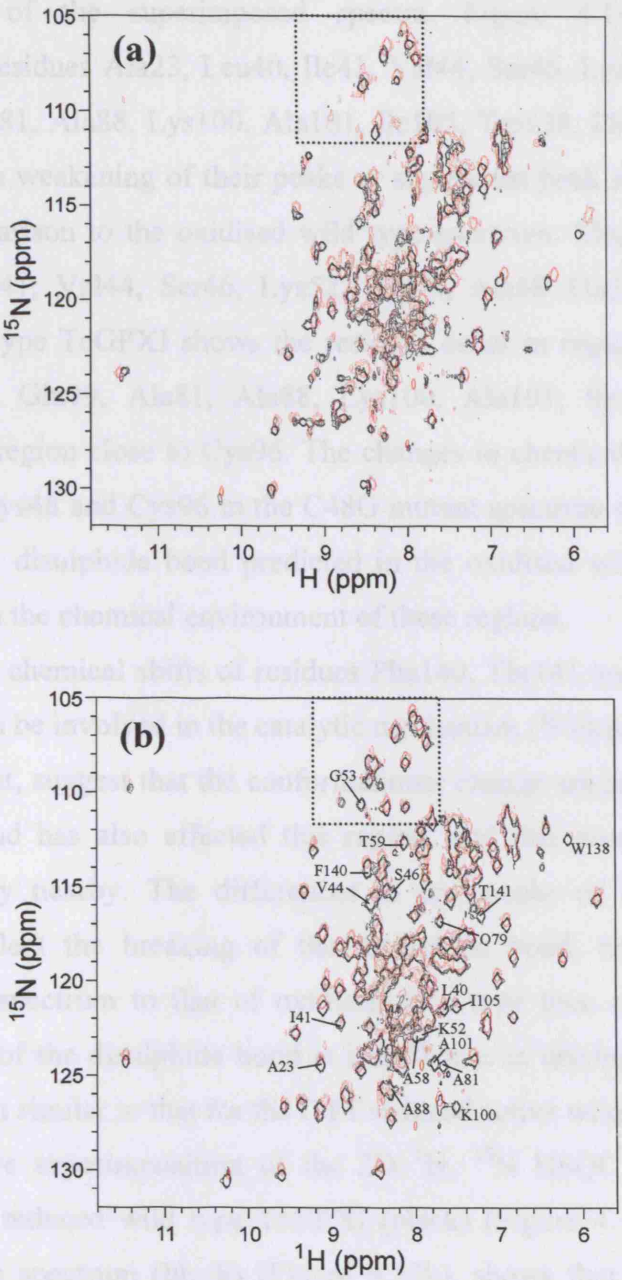


Figure 4.14: Superimposition of ^{1}H, ^{15}N HSQC (2D) spectra of C48G and wild type TcGPXI

Superimposition of 2D ^{1}H, ^{15}N HSQC spectra of ^{15}N-labelled C48G mutant TcGPXI (red) with (a) reduced wild type TcGPXI (black), and (b) oxidised wild type TcGPXI (black) in 50mM sodium phosphate, pH 6.5, 300mM NaCl.

The boxed area highlights a region where cross peaks in the C48G mutant spectrum are more similar to the oxidised spectrum than the reduced wild type spectrum. Residues showing peak intensity weakening or chemical shifts greater than 1 linewidth compared to the oxidised spectrum are indicated on spectrum (b).

peripheral areas of the superimposed spectra. Figure 4.14b shows peaks corresponding to residues Ala23, Leu40, Ile41, Val44, Ser46, Lys52, Gly53, Ala58 Thr59, Gln79, Ala81, Ala88, Lys100, Ala101, Ile105, Trp138, Phe140, and Thr141 have experienced a weakening of their peaks or significant peak shift greater than 1 linewidth in comparison to the oxidised wild type spectrum. Mapping the residues Ala23, Leu40, Ile41, Val44, Ser46, Lys52, Gly53, Ala58 Thr59 to the primary sequence of wild type TcGPXI shows the residues occur in regions around Cys48. Similarly residues Gln79, Ala81, Ala88, Lys100, Ala101, Ile105 occur in the primary sequence region close to Cys96. The changes in chemical shifts of residues localised around Cys48 and Cys96 in the C48G mutant spectrum are consistent with the absence of the disulphide bond predicted in the oxidised wild type sample as there is a change in the chemical environment of these regions.

Changes in chemical shifts of residues Phe140, Thr141 and Trp138 (Trp138 residue is known to be involved in the catalytic mechanism (Wilkinson et al., 2000a)) in the C48G mutant, suggest that the conformational change arising from the loss of the disulphide bond has also effected this region, and that possibly the catalytic Trp138 is spatially nearby. The differences in crosspeaks of the C48G mutant spectrum may reflect the breaking of the disulphide bond, however the closer agreement of the spectrum to that of oxidised wild type than reduced wild type, suggests that loss of the disulphide bond is insufficient in driving the protein fully into a conformation similar to that for the DTT reduced active wild type form.

Comparative superimposition of the 2D ^{1}H, ^{15}N HSQC spectra of C96G mutant (red) with reduced wild type TcGPXI (black) (Figure 4.15a), and with the oxidised wild type spectrum (black) (Figure 4.15b), shows that the C96G mutant spectrum is also more similar to that of the oxidised wild type ^{1}H, ^{15}N HSQC. The boxed area in Figure 4.15 highlights an example of a region where crosspeaks in the C96G mutant spectrum are more similar in number and position to that of the oxidised wild type spectrum than the reduced wild type spectrum.

Resonances that exhibit peak weakening or have shifted greater than one linewidth are highlighted in Figure 4.15b. Peaks corresponding to residues Val21, Val44, Ser46, Lys52, Gly54, Ala58, Lys100, Ala101, Asp102, Phe103 and Phe140 exhibit changes of peak weakening or shifts greater than 1 linewidth. These residues correlate to regions in the primary sequence around Cys48 and Cys96 and the

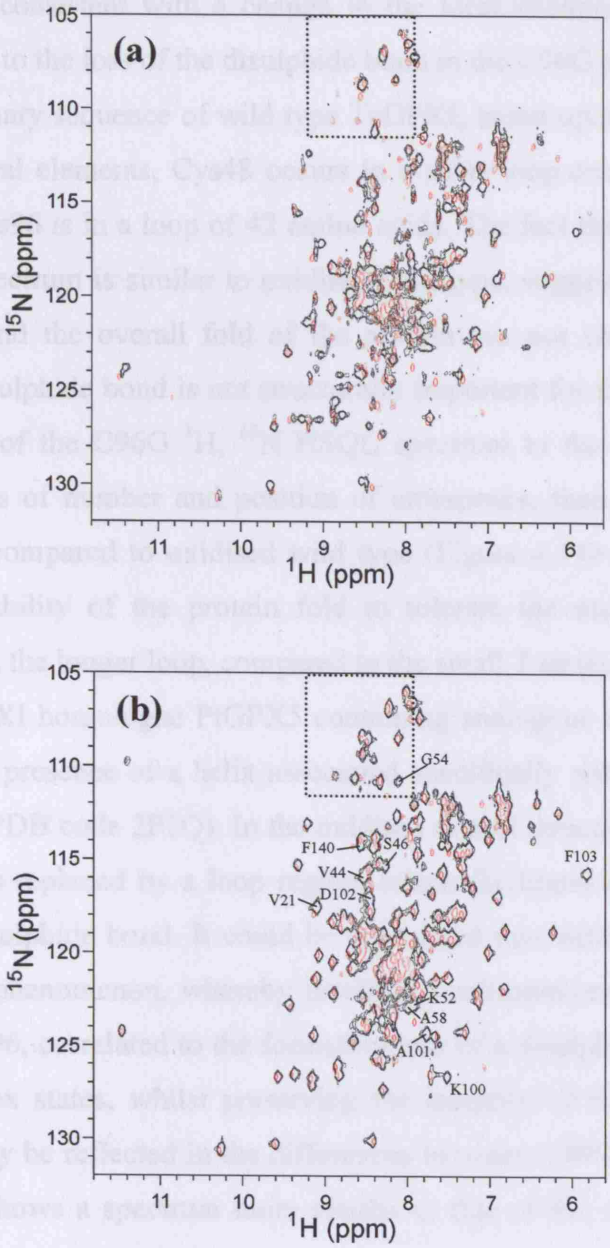


Figure 4.15: Superimposition of ^{1}H, ^{15}N HSQC (2D) spectra of C96G and wild type TcGPXI

Superimposition of 2D ^{1}H, ^{15}N HSQC spectra of ^{15}N-labelled C96G mutant TcGPXI (red) with (a) reduced wild type TcGPXI (black), and (b) oxidised wild type TcGPXI (black) in 50mM sodium phosphate, pH 6.5, 300mM NaCl.

The boxed area highlights a region where crosspeaks in the C96G mutant spectrum are more similar to the oxidised spectrum than the reduced wild type spectrum. Residues showing peak intensity weakening or chemical shifts greater than 1 linewidth compared to the oxidised spectrum are indicated on spectrum (b).

catalytic Trp138, consistent with a change in the local chemical environment for these residues due to the loss of the disulphide bond in the C96G mutant protein.

In the primary sequence of wild type TcGPXI, based upon the CSI predicted secondary structural elements, Cys48 occurs in a short loop consisting of 7 amino acids, whereas Cys96 is in a loop of 42 amino acids. The fact that the C96G mutant ^{1}H, ^{15}N HSQC spectrum is similar to oxidised wild type, suggests that upon loss of the disulphide bond the overall fold of the protein has not changed, which may reflect that the disulphide bond is not structurally important for the overall fold. The greater similarity of the C96G ^{1}H, ^{15}N HSQC spectrum to the oxidised wild type spectrum, in terms of number and position of crosspeaks, than the C48G ^{1}H, ^{15}N HSQC spectrum compared to oxidised wild type (Figure 4.14b and Figure 4.15b), may reflect the ability of the protein fold to tolerate the mutation and loss of disulphide bond in the longer loop, compared to the small 7 amino acid loop.

The TcGPXI homologue PtGPX5 containing analogous cysteine to TcGPXI Cys96, shows the presence of a helix associated specifically with the reduced form crystal structure (PDB code 2P5Q). In the oxidised crystal structure form (PDB code 2P5R) the helix is replaced by a loop region, which facilitates the formation of an intramolecular disulphide bond. It could be speculated that wild type TcGPXI may exhibit a similar phenomenon, whereby localised conformational changes occur in the region of Cys96, correlated to the formation/loss of a disulphide bond associated with specific redox states, whilst preserving the integrity of the remainder of the structure. This may be reflected in the differences between C48G and C96G spectra, where the latter shows a spectrum more similar to that of the stable oxidised wild type spectrum.

4.10 Ligand binding studies of wild type TcGPXI and mutants

NMR spectroscopy provides a set of residue-specific probes with which protein-ligand interactions may be investigated. Specifically monitoring the changes in crosspeaks in a 2D ^{1}H, ^{15}N HSQC spectrum, as a result of titrating a ligand into a protein sample can give sequence-specific information about residues that interact with the ligand (Clarkson and Campbell, 2003). Conformational changes of the protein may be induced through the binding of a ligand. Such conformational

changes may occur in small localised regions, where residues alter their local torsion angles to accommodate the ligand, or changes may occur in regions that lead to an overall change in the tertiary structure of the protein. The 2D ^{1}H, ^{15}N HSQC spectrum, which can contain a crosspeak for every amino acid in the primary sequence (except proline), is ideal for monitoring ligand titrations because a readout across the whole protein can be obtained in a single experiment. The addition of a ligand that binds to the protein will affect the chemical environment of residues that make contact with the ligand, resulting in a change in crosspeak position between free and ligand bound protein ^{1}H, ^{15}N HSQC spectra. This is observed as shifting or broadening of the cross-peaks in a 2D ^{1}H, ^{15}N HSQC due to ligand interaction (Wuthrich, 2002). In some cases the crosspeaks may even disappear due to the occurrence of intermediate exchange on the NMR timescale (Miller, 2000), or because the size of the complex is too large to be observable in the ^{1}H, ^{15}N HSQC spectrum.

Ligand-binding titrations using the known substrate t-butyl hydroperoxide (Wilkinson et al., 2000a) were carried out. Published work has shown that the reduced state is kinetically active in the reduced form. In order to assess the effects of substrate on the samples, a five fold excess of the substrate was added to the oxidised wild type TcGPXI sample, reduced wild type TcGPXI, C48G, C96G mutant proteins and compared to a ^{1}H, ^{15}N HSQC spectrum recorded without ligand.

The ^{1}H, ^{15}N HSQC of oxidised wild type TcGPXI with (red) and without (black) t-butyl hydroperoxide (Figure 4.16a), does not show any changes upon addition of the ligand, suggesting that the protein is required to be in a reduced redox state for the ligand to bind.

The ^{1}H, ^{15}N HSQC spectrum of reduced wild type TcGPXI is characterised by significantly less crosspeaks compared to that of the oxidised spectrum, which may be indicative of intermediate chemical exchange processes occurring once the disulphide bond has been broken. Figure 4.16b shows the ^{1}H, ^{15}N HSQC spectrum with (red) and without (black) the ligand titrated into reduced wild type TcGPXI. As can be seen there is no change in the spectrum suggesting that either the ligand is not binding to the sample of reduced wild type TcGPXI, or that the ligand is binding to residues whose crosspeaks are not present in the spectrum.

The C48G and C96G mutant protein samples were subjected to ligand

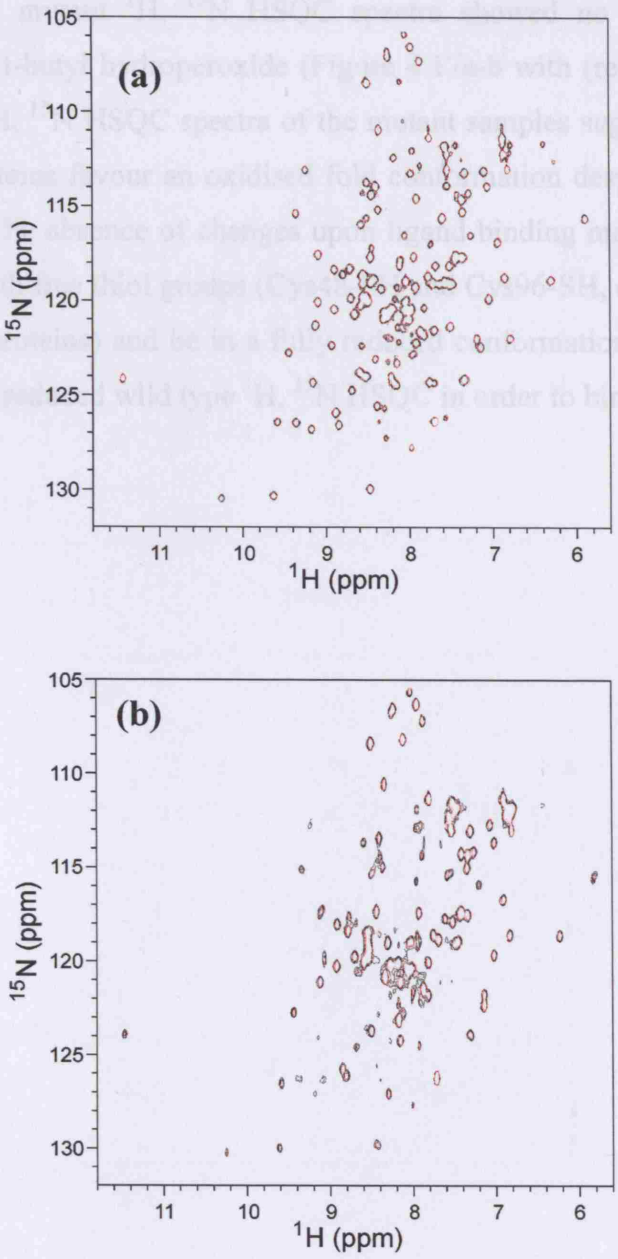


Figure 4.16: Ligand binding of t-butyl hydroperoxide to reduced and oxidised wild type TcGPXI

Superimposition of 2D ^{1}H, ^{15}N HSQC spectra of ^{15}N-labelled (a) oxidised wild type TcGPXI without (black) and with ligand (red) and (b) reduced wild type TcGPXI without (black) and with ligand (red) in 50mM sodium phosphate, pH 6.5, 300mM NaCl.

binding studies. Both mutant ^{1}H, ^{15}N HSQC spectra showed no changes upon addition of the ligand t-butyl hydroperoxide (Figure 4.17a-b with (red) and without (black) ligand). The ^{1}H, ^{15}N HSQC spectra of the mutant samples suggest the C48G and C96G mutant proteins favour an oxidised fold conformation despite the loss of the disulphide bond. The absence of changes upon ligand binding may indicate that the protein requires both free thiol groups (Cys48-SH and Cys96-SH, one of which is absent in the mutant proteins) and be in a fully reduced conformational fold similar to that observed in the reduced wild type ^{1}H, ^{15}N HSQC in order to bind ligand.

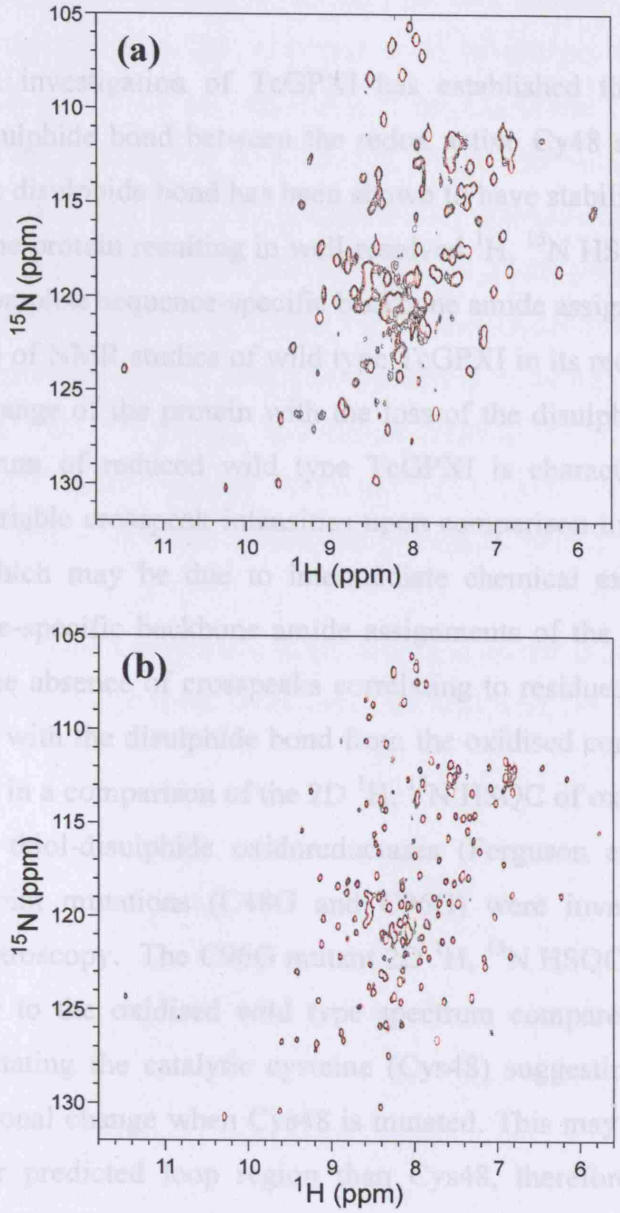


Figure 4.17: Ligand binding of t-butyl hydroperoxide to C48G and C96G mutant TcGPXI

Superimposition of 2D ^{1}H, ^{15}N HSQC spectra of ^{15}N-labelled (a) C48G mutant TcGPXI without (black) and with ligand (red) and (b) C96G TcGPXI without (black) and with ligand (red) in 50mM sodium phosphate, pH 6.5, 300mM NaCl.

4.11 Summary

The NMR investigation of TcGPXI has established the presence of an intramolecular disulphide bond between the redox active Cy48 and Cys96 of wild type TcGPXI. The disulphide bond has been shown to have stabilising effects on the conformation of the protein resulting in well-resolved ^{1}H, ^{15}N HSQC spectrum, thus facilitating near complete sequence-specific backbone amide assignments.

The results of NMR studies of wild type TcGPXI in its reduced form show a conformational change of the protein with the loss of the disulphide bond. The ^{1}H, ^{15}N HSQC spectrum of reduced wild type TcGPXI is characterised by missing crosspeaks and variable crosspeak intensities upon comparison to the oxidised wild type spectrum, which may be due to intermediate chemical exchange processes. Tentative sequence-specific backbone amide assignments of the reduced wild type spectrum reveal the absence of crosspeaks correlating to residues of predicted loop regions associated with the disulphide bond from the oxidised conformation. This is similarly observed in a comparison of the 2D ^{1}H, ^{15}N HSQC of oxidised and reduced SelM and Sep15 thiol-disulphide oxidoreductases (Ferguson et al., 2006). Two cysteine single point mutations (C48G and C96G) were investigated by NMR heteronuclear spectroscopy. The C96G mutant 2D ^{1}H, ^{15}N HSQC spectrum exhibits stronger similarity to the oxidised wild type spectrum compared to the spectrum obtained from mutating the catalytic cysteine (Cys48) suggesting the presence of greater conformational change when Cys48 is mutated. This may be because Cys96 occurs in a larger predicted loop region than Cys48, therefore the mutation of cysteine to glycine is more tolerated by the loop region and any resultant conformational changes do not affect the rest of the structure. The fact that both the C48G and C96G mutant ^{1}H, ^{15}N HSQC spectra are more similar to that of oxidised wild type TcGPXI suggests that the loss of a disulphide bond is insufficient in driving the mutant proteins into a reduced active conformational fold.

Ligand binding studies show that the wild type protein is unable to bind ligand in its oxidised state suggesting that the enzyme in a reduced active redox state is required for catalysis; whilst possible chemical exchange phenomena in the reduced form hinder the exact conclusion of those residues involved in ligand binding. It may be possible that the presence of an electron donor (redox shuttle protein *e.g.* tryparedoxin) is also required for ligand binding and so could enable

chemical exchange processes to be overcome, so that further NMR experiments could provide an insight into the structural state of the protein in its reduced conformational fold. The NMR investigations of the C48G and C96G mutants with the ligand showed no binding, but their ^{1}H, ^{15}N HSQC spectra are characterised as being more oxidised-like and so do not contain a fully reduced active conformational fold, that may be required for ligand binding. As well as the possibility of trying to produce a non-aggregating globular folded C77G mutant sample (similar to C48G and C96G mutant proteins), Cys77 could be mutated to an alternative residue, which could facilitate 2D ^{1}H, ^{15}N HSQC and ligand binding NMR experiments. Site directed mutagenesis could be used to mutate both Cys48 and Cys96 residues (double mutant) to assess whether the resultant protein adopts a DTT- reduced active fold by NMR.

The use of enzyme kinetics is a valuable tool to establish the catalytic activity of the mutants to suggest a possible mechanism for wild type TcGPXI.

4.12 Antioxidant system mechanisms

The following section describes the enzymes that are involved in antioxidant defence systems. Antioxidant defence plays a crucial role in rapidly growing and multiplying organisms. In trypanosoma, reactive oxygen species (ROS) occur as a result of endogenous reactions within an organism and as a product of the host immune response. The two major lines of antioxidant defence in mammalian organisms are the glutathione/glutathione reductase and thioredoxin/thioredoxin reductase systems (Figure 4.18). The dithiol components of the antioxidant systems and their numerous functions have been widely researched and shown to transfer reducing equivalents through recycling between the oxidised (S-S) and reduced redox states (-SH). Despite similarities between the two systems, it has been revealed that the glutathione/glutathione reductase and thioredoxin/thioredoxin reductase systems are exclusive from each other in their functions (Fairlamb and Cerami, 1992).

4.13 Glutathione/glutathione reductase and thioredoxin/thioredoxin reductase systems

In mammalian glutathione/glutathione reductase systems, glutathione-dependent peroxidases have been revealed to reduce hydroperoxides at the expense of glutathione (GSH). The glutathione/glutathione reductase system has been well researched and characterised where the crystal structure of human glutathione reductase has been solved (PDB code 1GRA). In brief, the overall mechanism involves the transfer of electrons to NADPH-dependent flavin-containing glutathione reductase (GR), whereby this enzyme is reduced. Subsequently, the enzyme is oxidised upon electron transfer to reduce oxidized glutathione (generating two molecules of reduced glutathione, GSH) (Figure 4.18a) (Turrens, 2004). A similar process is observed in thioredoxin/thioredoxin reductase systems (Figure 4.18b). Electrons are transferred from NADPH via FAD to the active site disulphide of thioredoxin reductase (TrxR), which then reduces the redox-active substrate protein thioredoxin (Trx) (Mustacich and Powis, 2000). Thioredoxins are characterised by a conserved active site W-C-G-P-C-K sequence motif (Holmgren, 1985) and has many

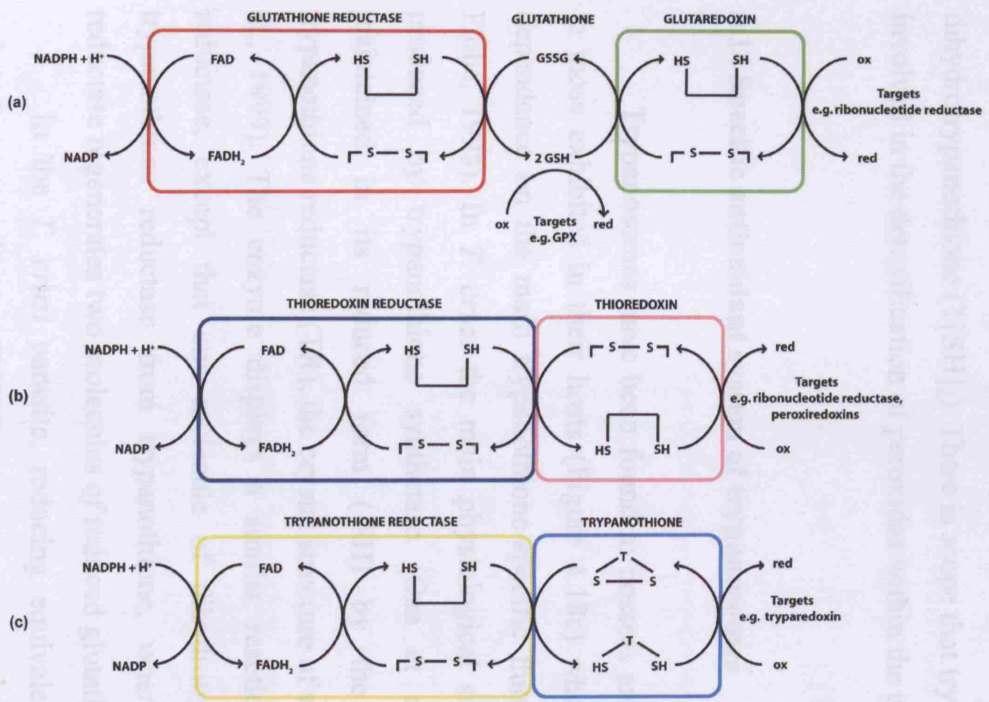


Figure 4.18: Antioxidant defence systems in mammalian hosts and trypanosome parasites

(a) Glutathione/glutathione reductase system. Reducing equivalents are transferred from NADPH to, glutathione (green) via glutathione reductase (red) and then to glutaredoxin/glutathione peroxidase (GPX) for reduction of targets such as ribonucleotide reductase or hydroperoxides, respectively.

(b) Thioredoxin/thioredoxin reductase system. Reducing equivalents are transferred from NADPH to thioredoxin reductase (blue) and thioredoxin (pink) for reduction of targets such as peroxiredoxins or ribonucleotide reductase.

(c) Trypanothione/trypanothione reductase system. Reducing equivalents are transferred from NADPH to trypanothione (cyan) via trypanothione reductase (yellow) for reduction of targets such as tryparedoxin which in turn reduces hydroperoxides. (Taken from Fairlamb and Cerami, 1992).

functions including its ability to act as a redox shuttle providing electrons to thioredoxin-peroxidases (or peroxiredoxins), which in turn reduce hydroperoxides.

Until recently it was thought that thioredoxins did not exist in trypanosomes. However, research has revealed the existence of thioredoxin redox molecules in both *T. brucei* and *T. cruzi* (Piattoni et al., 2006) but so far no evidence for the presence of a thioredoxin reductase enzyme has been found (Muller et al., 2003). Interestingly, *T. brucei* thioredoxin has been shown to be able to transfer reducing equivalents to a trypanosome-specific tryparedoxin peroxidase and be reduced by redox thiol dihydrotrypanothione ($T[SH]_2$). There is scope that trypanosomal thioredoxin may be involved in the detoxification of peroxides within the parasite (Piattoni et al., 2006).

4.14 Specific antioxidant system of trypanosomes

Trypanosomes have been found to possess an analogous antioxidant system to those exhibited in their hosts (Figure 4.18c) where the main difference is the dependence on the main trypanothione-specific thiol trypanothione (Kelly, 2003; Flohe, 1998). In *T. cruzi* the main physiological substrate is homotrypanothione produced by trypanothione synthetase (Oza et al., 2002). Trypanothione is maintained in its reduced form (-SH) by the flavoprotein oxidoreductase trypanothione reductase (TR), the crystal structure of which has been solved (Bond et al., 1999). The enzyme displays a similar reaction mechanism to glutathione reductase, except that one molecule of dihydrotrypanothione is generated by trypanothione reductase from trypanothione, whereas contrastingly glutathione reductase regenerates two molecules of reduced glutathione.

In the *T. cruzi* parasite, reducing equivalents may be transferred from reduced trypanothione ($T[SH]_2$) to glutathione (GSH), ascorbate (ASC) or tryparedoxin (TPX) (Figure 4.19). Despite the presence of glutathione in *T. cruzi*, glutathione reductase is not found in trypanosome parasites. The level of free reduced glutathione in *T. cruzi* has been found to be quite low in comparison to mammalian hosts (Raoult, 1993; Bond et al., 1999).

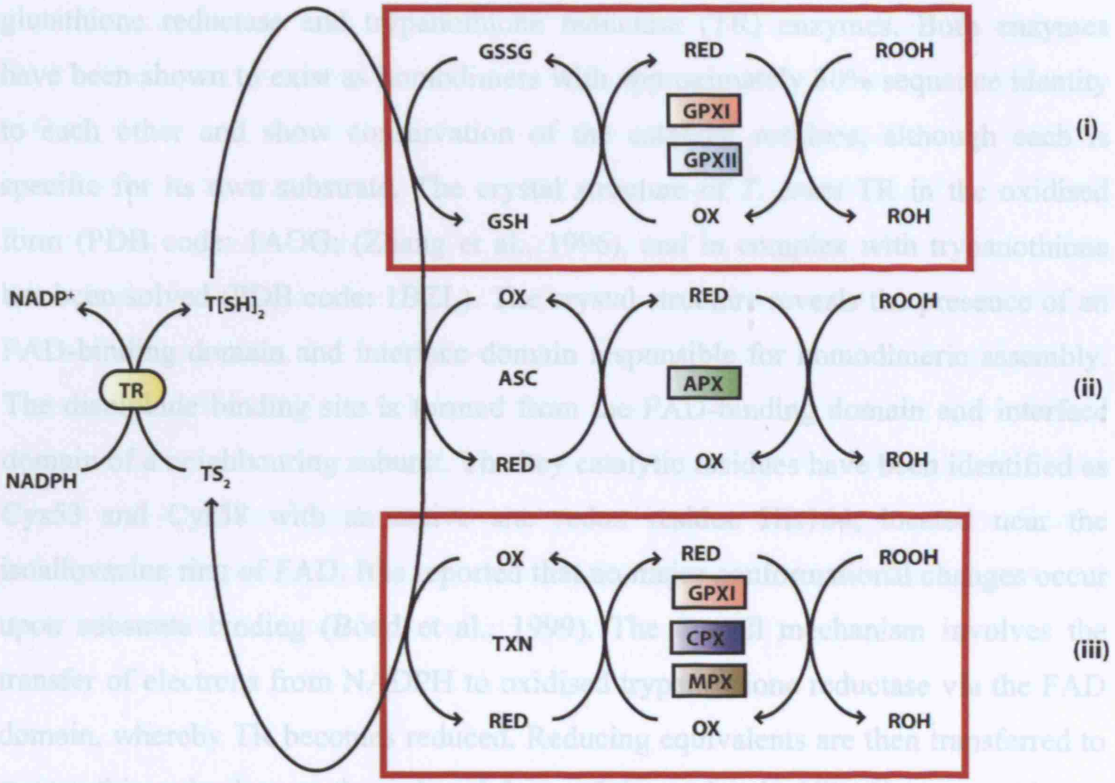


Figure 4.19: Antioxidant pathway in *Trypanosoma cruzi*

Reducing equivalents are transferred from NADPH to trypanothione, where trypanothione reductase (TR) is used to generate dihydrotrypanothione ($T[SH]_2$). Reducing equivalents can then be transferred to:

i) glutathione (GSH) and glutathione peroxidase enzymes (TcGPXI, TcGPXII);

ii) ascorbate (ASC) and ascorbate peroxidase (APX);

iii) tryparedoxin (TXN) and glutathione peroxidase (TcGPXI or peroxiredoxin enzymes (cytosolic, CPX; mitochondrial, MPX)

The final stage is the reduction of hydroperoxide species.

4.15 Trypanothione reductase (TR)

Trypanothione reductase is essential in maintaining reduced trypanothione as part of the oxidoreductase cascade. Many similarities have been established between glutathione reductase and trypanothione reductase (TR) enzymes. Both enzymes have been shown to exist as homodimers with approximately 30% sequence identity to each other and show conservation of the catalytic residues, although each is specific for its own substrate. The crystal structure of *T. cruzi* TR in the oxidised form (PDB code: 1AOG; (Zhang et al., 1996), and in complex with trypanothione has been solved (PDB code: 1BZL). The crystal structure reveals the presence of an FAD-binding domain and interface domain responsible for homodimeric assembly. The disulphide binding site is formed from the FAD-binding domain and interface domain of a neighbouring subunit. The key catalytic residues have been identified as Cys53 and Cys58 with an active site redox residue His164, located near the isoalloxazine ring of FAD. It is reported that no major conformational changes occur upon substrate binding (Bond et al., 1999). The overall mechanism involves the transfer of electrons from NADPH to oxidised trypanothione reductase via the FAD domain, whereby TR becomes reduced. Reducing equivalents are then transferred to trypanothione leading to the reduced form of the molecule (dihydrotrypanothione). Trypanothione reductase has been considered as a good drug target due to its relatively upstream functional role in the trypanosomal antioxidant system. However, it has been suggested that *T. cruzi* parasite may survive even when TR activity is reduced by 90% making this role questionable (Wilkinson et al., 2002a).

4.16 Tryparedoxin

T. cruzi tryparedoxin (TXN) is a thiol-disulphide oxidoreductase part of the trypanothione-linked antioxidant pathway, acting as a redox shuttle between trypanothione and tryparedoxin-peroxidases or glutathione peroxidase enzymes. This molecule utilises the redox properties of disulphide bonding. The crystal structure of *Crithidia fasciculata* tryparedoxin reveals that the redox molecule is a homologue of thioredoxin. Tryparedoxin contains a conserved catalytic site sequence motif ("WCPPC") distinct from the analogous catalytic sequence motif observed in

thioredoxin ("WCGPCK") (Lopez et al., 2000; Gommel et al., 1997). Despite this similarity in catalytic site motif and overall fold (both exhibit a thioredoxin fold), tryparedoxin is somewhat larger than thioredoxin with extra secondary structural regions (Alphey et al., 2003). In *C. fasciculata* tryparedoxin, the redox active disulphide involves residues Cys40 and Cys43, positioned under the indole ring of Trp39. The crystal structure of tryparedoxin in *T. brucei* has been solved (PDB code: 1O73; (Alphey et al., 2003)). The active site residues and neighbouring residues (considered to be associated to the binding of trypanothione) are highly conserved. The presence of a conformationally adaptable tryptophan residue, also observed in thioredoxin, is thought to contribute to the interactions that allow for a transfer of reducing equivalents. In *T. brucei* it has been found that various isoforms of tryparedoxin exist differing their cellular localisation and that the presence of tryparedoxin is essential for oxidative defence (Comini et al., 2007). Studies have shown that *T. brucei* tryparedoxin may not be a suitable drug target as reduction of activity would have to have to be lowered to less than 5% for a continuous period of time to be a potential threat to parasites (Comini et al., 2007).

4.17 Tryparedoxin-peroxidase enzymes

In the final stage of the *T. cruzi* antioxidant defence pathway, hydroperoxides are reduced by tryparedoxin peroxidases (peroxiredoxins). The crystal structure of tryparedoxin peroxidases from *T. cruzi* and *C. fasciculata* has been solved (Alphey et al., 2000; Pineyro et al., 2005). Tryparedoxin peroxidases belong to the peroxiredoxin (Prx) family of proteins. In *T. cruzi* cytosolic (TcCPX) and mitochondrial (TcMPX) forms have been shown to belong to the typical 2-Cys mechanistic subgroup of the peroxiredoxin family (Wood et al., 2003; Pineyro et al., 2005). The dimerisation of mitochondrial and cytosolic peroxiredoxin isoforms, is facilitated by the formation of intermolecular disulphide bonds (Wilkinson et al., 2000b). The crystal structure of TcCPX in its reduced active form shows the monomer exhibits a thioredoxin fold, which participates in a pentameric organisation of symmetrical homodimers (decameric). The structure is shown to be more similar to the structure of human thioredoxin-peroxidase than to the structure of *C. fasciculata* tryparedoxin peroxidase despite a higher sequence similarity. The active

site is located in a pocket that opens to the dimer-dimer interface close to the peroxidatic cysteine. The resolving cysteine is from a second molecule of a homodimer located in a hydrophobic environment, considered essential for catalytic activity. In the reduced form the two redox active cysteine residues are considered too distant to form a disulphide bond, therefore it has been suggested that there is a conformational change after the first stage of catalysis (whereby sulfenic acid formed upon reduction of the hydroperoxide substrate), thus facilitating the two cysteine residues to move into close proximity to create a disulphide bridge as an oxidised intermediate. Residues Phe50 and Thr49 (Phe50 is shown to be conserved in other typical 2-Cys peroxiredoxins but variable in atypical 2-Cys peroxiredoxins) are postulated to be involved in facilitating structural changes (Pineyro et al., 2005). Tryparedoxin peroxidases catalyse the reduction of various hydroperoxides significantly faster than the spontaneous reactions of the hydroperoxides with dihydrotrypanothione. This suggests that it is essential that the trypanosomal peroxiredoxins are kept in a reduced state by the redox shuttle tryparedoxin, so that sensitive cellular targets are not exposed to harmful hydroperoxides for sustained periods of time. The TcCPX and TcMPX peroxiredoxin enzymes have been shown to metabolize hydrogen peroxide, as does the haem-binding ascorbate peroxidase (APX) (Wilkinson et al., 2002c; Wilkinson et al., 2000b).

4.18 *T. cruzi* glutathione-dependent type peroxidases (GPX)

The *T. cruzi* parasite, in addition to the tryparedoxin-linked peroxiredoxin enzymes, contains glutathione-dependent peroxidase enzymes TcGPXI and TcGPXII. Both enzymes are monomeric and are unable to metabolise hydrogen peroxide. Their substrate specificity of both enzymes is restricted to fatty acid and phospholipid hydroperoxides, where TcGPXI is additionally able to metabolise short chain hydroperoxides (Wilkinson et al., 2000a). TcGPXI has been established to function both in a glutathione-dependent and tryparedoxin-dependent manner, although the latter exhibits the greater turnover (Wilkinson and Kelly, 2003). TcGPXII, occurs in the endoplasmic reticulum and displays ping pong kinetics only utilising glutathione as an electron donor. The reaction is limited by the glutathione-dependent reduction of the enzyme similar to TcGPXI (Wilkinson et al., 2002b).

Oxidised glutathione resulting from the reduction of the peroxidase enzymes can be recycled into the reduced form by interaction with trypanothione (non-enzymatically or via a thiol transferase).

4.19 Biochemical characterisation of TcGPXI using kinetic analyses

Published kinetic analyses using TcGPXI with a variety of substrates has revealed the enzymes ability to use both glutathione and tryparedoxin as electron donors. The enzyme exhibits substrate specificity similar to TcGPXII in that both TcGPXI and TcGPXII can metabolise fatty acid and phospholipid hydroperoxides, but neither can metabolise hydrogen peroxide. The difference in substrate specificities between TcGPXI and TcGPXII, is that TcGPXI can metabolise short chain alkyl hydroperoxides e.g. cumene hydroperoxide, t-butyl hydroperoxide whereas TcGPXII cannot. In the glutathione-mediated pathway, TcGPXI displays ordered kinetics although this is speculative, as accurate kinetic investigation has been proven to be difficult due to the low affinity of the enzyme to GSH (K_m >5mM) and high affinity to the hydroperoxide substrate (Wilkinson and Kelly, 2003). The activity of TcGPXI in the glutathione-mediated pathway has been investigated with a variety of substrates. In particular, the published apparent K_m and V_{max} values obtained for t-butyl hydroperoxide substrate are 5.64 ± 0.53 μM and 33.66 ± 0.89 nmol of NADPH oxidised min^{-1} mg of protein $^{-1}$, respectively.

TcGPXI is unable to use trypanothione as a direct electron donor and like its trypansomal peroxiredoxin counterparts, additionally requires the redox shuttle tryparedoxin. TcGPXI has been shown to have a greater activity with tryparedoxin than for glutathione, where metabolism of hydroperoxides is 8-15 fold higher for the former than that of the latter, via the glutathione-linked pathway. All reactions in the tryparedoxin-linked pathway exhibit a ping pong mechanism where the enzymes function sequentially. Published kinetic investigation using the substrate cumene hydroperoxide at a range of concentrations (2-20 μM), has revealed apparent V_{max} and K_m values of 1899 ± 155 nmol NADPH oxidized min^{-1} mg^{-1} and 16.1 ± 1.0 μM, respectively (Wilkinson et al., 2002a). The rate-limiting step has been found to be the reduction of tryparedoxin by trypanothione. The use of different electron donors has

been suggested to reflect the parasites use of different electron donors according to the cellular localisation.

4.20 Use of mutational studies to elucidate the catalytic mechanism

The use of mutational analysis in order to elucidate the catalytic mechanism of glutathione-dependent peroxidase enzymes has been well established. A good example has been performed in classic mammalian selenium-containing phospholipid hydroperoxides. Mutational analysis facilitated the identification of the conserved catalytic triad selenocysteine (or cysteine), glutamine and tryptophan in the mammalian SeCys-PHGPX enzyme (Maiorino et al., 1998). The first stage in the reaction is oxidation of the thiol group of PHGPX by hydroperoxide yielding a selenic acid derivative. The oxidised enzyme undergoes reaction with a single molecule of glutathione, resulting in an intermediate in which the sulphur of the glutathione molecule is bonded to the selenium moiety via a selenodisulphide bridge. Finally, the reduced enzyme is regenerated by a second molecule of glutathione, which eliminates the selenodisulphide bridge, thus releasing oxidised glutathione (GSSG) (Maiorino et al., 1998). In mammals this is recycled into the reduced form by GR. Mutational analysis has facilitated the key catalytic residues involved in the reaction described above to be deciphered.

Recently, the glutathione-dependent type enzyme in *T. brucei* (TbPxIII homologue) has been subjected to similar mutational analysis utilizing the tryparedoxin-dependent antioxidant pathway, and has identified the requirement of the conserved GPX-family catalytic triad (cysteine (Cys47), glutamine (Gln82) and tryptophan (Trp137)) and a resolving cysteine (Cys95) for enzyme activity. The results reported have illustrated that the *T. brucei* PxIII enzyme exhibits a mechanism characteristic of an atypical 2-Cys peroxiredoxin mechanism (Schlecker et al., 2007). The function of the third cysteine residue in *T. brucei* Cys 76 (analogous to *T. cruzi* Cys77) is unknown, however Jung et al., 2002 have reported and suggested a structural role for the analogous cysteine in an enzyme from Chinese cabbage.

4.21 Use of the glutathione-linked and tryparedoxin-linked assays to investigate the postulated reaction mechanism of wild type TcGPXI and its associated cysteine mutated enzymes (C48G, C96G and C77G)

In order to investigate the catalytic activity of wild type TcGPXI and of the TcGPXI cysteine mutations (C48G, C96G and C77G), enzyme assays were carried out using previously published and established protocols (Wilkinson et al., 2000a; Wilkinson et al., 2002a).

4.21.1 Catalytic investigation of wild type and cysteine mutant TcGPXI enzymes using glutathione-linked assay

In order to carry out enzyme assays, fresh GPXI protein was expressed and purified as described in Chapter 2. The enzymatic reaction was set up as described in Section 2.18.

Briefly, as illustrated in reaction (i) of Figure 4.19, components of the enzymatic assay were set up such that a constant supply of reduced glutathione was available by the presence of glutathione reductase such that no limiting factors affected the reaction in the coupled assay. Initially assay conditions were tested and established for the wild type enzyme using the substrate t-butyl hydroperoxide. In order to ensure Michaelis-Menten conditions of [S]>[E] were satisfied and that initial rates were observed, a 5-fold excess of substrate over total enzyme was used. Initial rates reaction was measured spectrophotometrically by the linear decrease in absorption at 340nm due to the oxidation of NADPH.

For the determination of kinetic parameters the concentration of t-butyl hydroperoxide substrate was varied across the published K_m value of 5.64 μM (utilising a range 3.6 - 50 μM of t-butyl hydroperoxide). Assays with the TcGPXI mutant enzymes (C48G, C96G and C77G) were carried out in the same manner as for the wild type TcGPXI enzyme. In all cases the background activity was negligible.

The reduced wild type TcGPXI enzyme was shown to exhibit activity in glutathione-mediated assays. Plotting the activity (nmols $min^{-1}mg^{-1}$) versus substrate concentration ($[S]_0$) yields a graph depicting a rectangular hyperbola which is usually

observed for enzymes that contain a single catalytic site and to which an adsorptive complex can be formed as is shown in the simple equation below:

$$\textbf{Equation (i)} \quad E + S \rightleftharpoons ES \rightleftharpoons E + P$$

(Where E is free enzyme, S is free substrate, ES is adsorptive complex and P is product).

The data was fitted to the Michaelis-Menten equation (Figure 4.20a), which takes the form of a rectangular hyperbola using SigmaPlot version 10.0:

$$\textbf{Equation (ii)} \quad v_i = (V_{max}\,[S]_0) / (K_m + [S]_0)$$

In order to confirm accurate values for V_{max} and K_m (*i.e.* those deduced from fitting the graph to a rectangular hyperbolas) a linear transformation was applied to generate an Eadie-Hofstee plot given by equation:

$$\textbf{Equation (iii)} \quad v = -K_m\,(v/[S]_0) + V_{max}$$

i.e. a graphical plot of activity (initial rates) vs. activity/$[S]_0$ will give apparent values for K_m (slope), V_{max} (y-intercept) and V_{max}/K_m (x-intercept). This plot is regarded as a more sensitive plot than the more commonly used Lineweaver-Burke plot, in that the activity/$[S]_0$ ratio is sensitive to change and thus any deviation from hyperbolic kinetics is detected.

The Eadie-Hofstee plot for the glutathione-linked assay is illustrated in Figure 4.20b. The linearity of the plot supports the suggestion that this enzyme shows single site saturation and adopts Michaelis-Menten behaviour. The Eadie-Hofstee plot provides apparent values of V_{max} (y intercept), $-K_m$ (slope) and V_{max}/K_m for the reaction of wild type TcGPXI with the substrate t-butyl hydroperoxide of 359 ± 19.4 nmol min^{-1} mg^{-1}, 52.1 ± 4.165 μM and 6.9 nmol min^{-1} mg^{-1} μM^{-1}respectively.

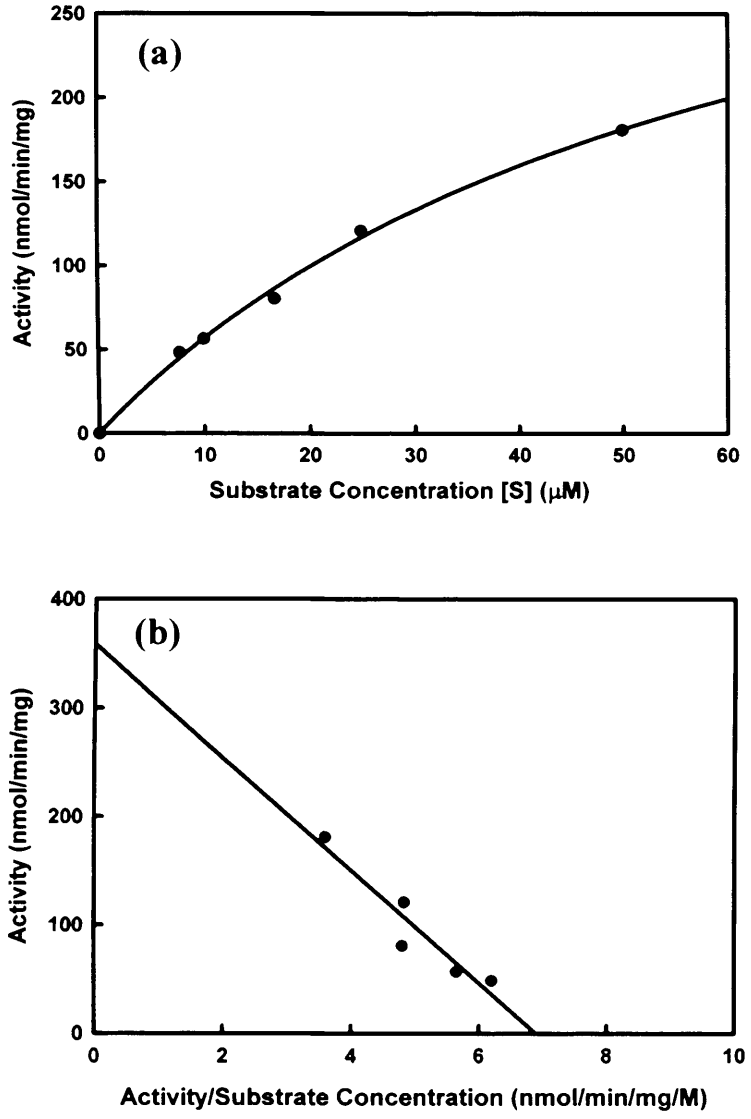


Figure 4.20: Kinetic investigation of wild type TcGPXI using the glutathione-linked antioxidant pathway

(a) Michaelis-Menten plot for wild type TcGPXI over a range of t-butyl hydroperoxide concentrations (0-50 μM) at a fixed amount of enzyme (2.2 μM). The data is fitted to a rectangular hyperbola as in Equation ii and the graph constructed in SIGMAPLOT 10.0 (Systat).

(b) Eadie-Hoftsee plot constructed from values obtained for the kinetic investigation of wild type TcGPXI using substrate concentrations spanning published K_m. The graph is linear and provides values of $-K_m$ (slope), V_{max} (extrapolation to the y ordinate) and efficiency constant, V_{max}/K_m, (extrapolation to the x ordinate). The apparent K_m, V_{max}, and V_{max}/K_m values are 52.1 ± 4.16 μM, 359 ± 19.4 nmol min^{-1} mg^{-1} and 6.9 nmol min^{-1} mg^{-1} μM^{-1}.

The values of K_m and V_{max} (from Figure 4.20b) are provided, albeit with some error due to extrapolation to both the x and y ordinate. Inspection of the published data and that repeated here reveals an approximately ten-fold difference in the values of V_{max} and K_m. Wilkinson *et al* report apparent values for K_m and V_{max} for full-length TcGPXI of 5.64 ± 0.53 μM and 33.66 ± 0.89 nmols min^{-1} mg^{-1} respectively. This 10-fold difference in activity might be accounted for by a difference in active site content. Additionally, the difference in values of K_m observed in this study and that of the published data from Wilkinson and colleagues might be attributed to differences in storage of the enzyme, where the enzyme used to generate the published data was less active due to oxidation of the dithiol catalytic components compared to the enzyme utilised in this investigation, which was made fresh. This oxidation, which may be irreversible might cause the GPXI protein in the Wilkinson et al paper to exhibit a different binding mode to the substrate, t-butyl hydroperoxide, thus trapping the substrate and having an effect on the apparent value of K_m (5.64 uM). The effective binding of the GPXI protein to the t-butyl hydroperoxide substrate (K_m 5.64 uM) as reported by Wilkinson and colleagues does not yield any advantage in catalysis as illustrated by the published value of V_{max}, 33.7 nmol/min/mg compared to 359 nmol/min/mg reported in this investigation (Table 4.1) - an approximate 10-fold difference in values. This 10-fold difference in values of V_{max} suggests a low active centre content attributed through possible irreversible oxidation may render the enzyme inactive thus lowering values of V_{max}. It is known to be a common problem that active centre content cannot be measured directly unless a specific titrant is available, which reacts with catalytic residues in the active centre in a specific manner. Thus it is most common to use total protein content as total enzyme concentration (László Polgár, 1989). A specific active site titrant such as 2,2-dipyridyl disulphide (2PDS) may be used to monitor active site thiol groups (Hussain et al., 2002). Correction for active centre content can be performed by making use of the efficiency constant V_{max}/K_m. The efficiency constant V_{max}/K_m (nmol min^{-1} mg^{-1} μM^{-1}) is 6.9 from this study and 6.0 from data published by Wilkinson and colleagues. The close agreement of the V_{max}/K_m ratio reported here (6.9) and that published (6.0) suggests that active centre content might well account for the 10-fold difference observed of the Michaelis parameters, V_{max} and K_m.

An investigation of the TcGPXI C48G, C77G and C96G mutant enzymes using

	K_m (μM)	V_{max} (nmol min^{-1} mg^{-1})	V_{max}/K_m (nmol min^{-1} mg^{-1} μM^{-1})
Published data* (Wilkinson et al., 2000a)	5.64 ± 0.53	33.68 ± 0.89	6.0
Michaelis-Menten plot*	61.2 ± 8.6	403 ± 37.2	6.6
Eadie-Hoftsee plot*	52.1 ± 4.16	359 ± 19.4	6.9
Michaelis-Menten plot**	45.6 ± 7.7	1255 ± 98.3	27.5
Eadie-Hoftsee plot**	50.7 ± 9.5	1317 ± 159.2	25.9

* Glutathione-mediated assay using t-butyl hydroperoxide substrate.

** Tryparedoxin-mediated assay using t-butyl hydroperoxide substrate.

Table 4.1: Kinetic parameters for TcGPXI

Comparison of wild type TcGPXI kinetic parameters K_m, V_{max} and V_{max}/K_m from tryparedoxin and glutathione-mediated antioxidant pathways with published data.

the glutathione coupled assay revealed an absence of catalytic activity. This might suggest that residues Cys48 and Cys96 are essential for catalysis. Furthermore, the activity observed from the assays of the TcGPXI cysteine mutant enzymes was found to be comparable to background activity, further reinforcing the requirement of these cysteine residues in catalysis and illustrating enzyme inactivity. The C77G mutant enzyme exhibited a lack of a detectable catalytic activity, which may be due to the fact that mutating the Cys77 residue to glycine, causes a significant structural change thereby rendering the enzyme inactive. This possible structural change is supported by the results of 1D ^{1}H NMR carried out on the TcGPXI C77G mutant enzyme, revealing the presence of unfolded components, compared to the spectra of TcGPXI wild type, C48G and C96G mutant enzymes. Alternatively, it may be possible Cys77 has a regulatory role in catalysis. The possibility for a regulatory role has been suggested in the homologue *T. brucei* PxIII that contains an analogous cysteine in addition to the proposed peroxidatic and resolving cysteine residues (Schlecker et al., 2007). Subsequent to the data presented in this investigation, work has been carried out in our laboratory where Cys77 was mutated to residues alanine and serine and glutathione-linked enzyme assays were repeated (Alaba 2008 unpublished). Catalytic activity was measured using the glutathione-mediated assay and revealed the presence of catalytic activity despite observations in protein instability. This further provides evidence strongly supporting the fact that Cys48 and Cys96 (postulated peroxidatic and resolving cysteine residues, respectively) indeed redox active residues required for substrate metabolism by the TcGPXI enzyme.

4.21.2 Catalytic investigation of wild type and cysteine mutant TcGPXI enzymes using the tryparedoxin-linked assay

It has been previously reported that TcGPXI is approximately 8-15 fold more catalytically active when coupled to the tryparedoxin linked pathway of the trypanosomal antioxidant system, than when coupled to the glutathione mediated pathway. Having utilised the glutathione-coupled assay to assess the catalytic activity of TcGPXI mutant enzymes, it was appropriate to use a more sensitive coupled system in order to validate the fact that the TcGPXI cysteine mutant enzymes were found to be catalytically inactive. Published data using the substrate cumene

hydroperoxide (substrate range 2-20 μM) suggests the apparent values of V_{max} and K_m as 1899 ± 155 nmol NADPH oxidized min^{-1} mg^{-1} and 16.1 ± 1.0 μM respectively. Kinetic parameters for the reduced wild type and cysteine mutant TcGPXI enzymes were investigated in a similar fashion as the published protocol but using the substrate t-butyl hydroperoxide, coherent with investigations carried out utilising the glutathione couple assay in this study, at a substrate range of 0-100 μM. Trypanothione reductase (TR) and tryparedoxin were prepared as described in Section 2.18.

In a similar method to the glutathione-linked assays, the oxidation of NADPH at 340nm was monitored using the substrate t-butyl hydroperoxide. Fresh TcGPXI wild type and TcGPXI cysteine mutant (C48G, C96G, C77G) enzymes were prepared as previously described (Chapter 2). The enzymatic reaction was initiated by the addition of TcGPXI (wild type or mutant TcGPXI) enzyme at a final concentration of 2 μM in a total assay volume of 1 ml at 30 °C, and the reaction subsequently followed by the decrease in absorbance of NADPH at 340 nm. Background rates were found to be negligible. The wild type TcGPXI enzyme was found to exhibit catalytic activity, and the resultant data was fitted to the Michaelis-Menten equation (Equation ii) taking form of a rectangular hyperbola (Figure 4.21a). In order to deduce apparent V_{max} and K_m values, the data was linearly transformed into the Eadie-Hofstee equation (Equation iii) and graphically plotted (Figure 4.21b). The apparent V_{max} and K_m values obtained from the Eadie-Hofstee plot are 1316 ± 159.2 nmol min^{-1} mg^{-1} and 50.7 ± 9.5 μM.

A useful measure of the efficiency of enzyme catalysis between the two types of coupled assays used in this study (i.e. tryparedoxin *vs.* glutathione) can be made by accounting for the efficiency constant V_{max}/K_m. The values in Table 4.1 reveal that for the glutathione coupled reaction, a V_{max}/K_m value of 6.9 is obtained in comparison to a V_{max}/K_m value of 25.9 for the tryparedoxin coupled assay, an approximately 3.5 fold difference. This would imply, and coherent with published data, that the tryparedoxin coupled assay is indeed more sensitive than that of the assay coupled with glutathione. Having established that the wild type TcGPXI enzyme was catalytically active, the TcGPXI mutant enzymes (C48G, C77G and C96G) were assessed in a similar fashion as the wild type enzyme. Unsurprisingly, the TcGPXI mutants were found to be inactive, thus reinforcing the results obtained with the

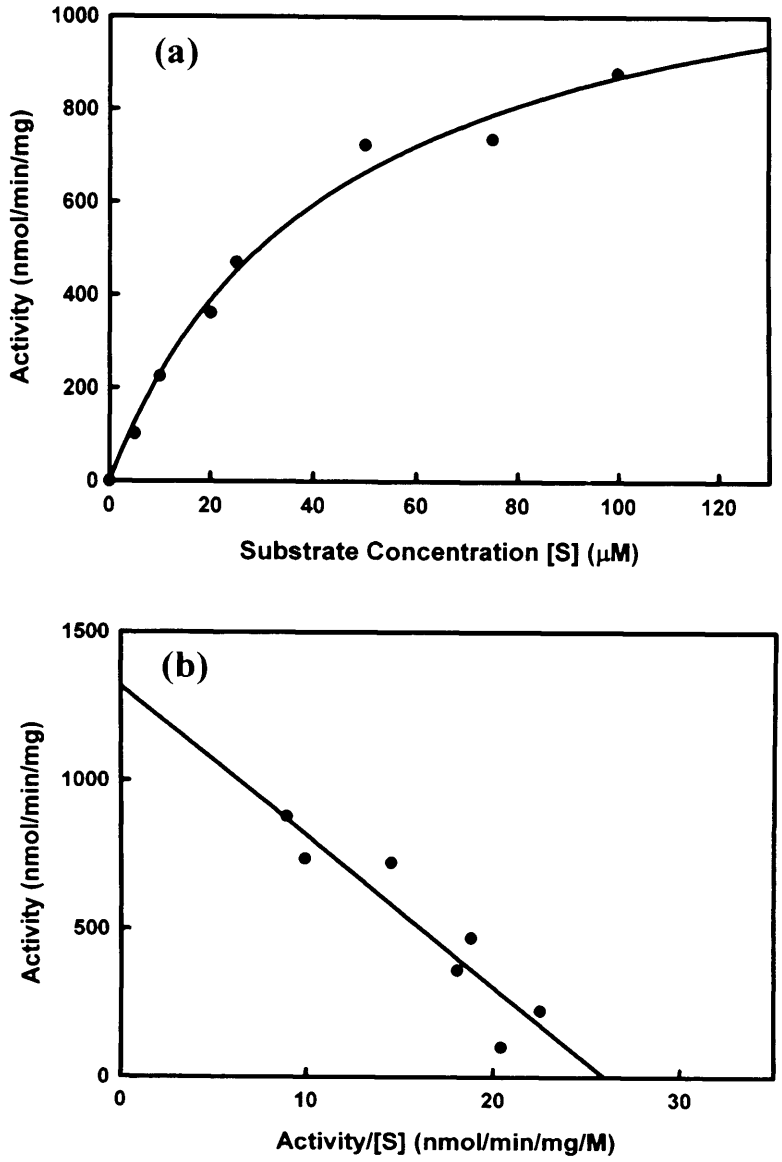


Figure 4.21: Kinetic investigation of wild type TcGPXI using the tryparedoxin-linked antioxidant pathway

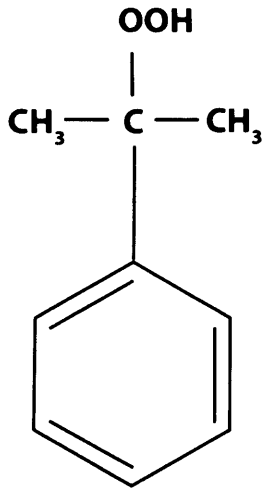
(a) Michaelis-Menten plot for wild type TcGPXI over a range of t-butyl hydroperoxide concentrations (0-100 μM) at a fixed enzyme concentration (2 μM). The data is fitted to a rectangular hyperbola as in Equation ii and the graph constructed in SIGMAPLOT 10.0 (Systat).

(b) Eadie-Hoftsee plot constructed from values obtained for the kinetic investigation of wild type TcGPXI using substrate concentrations (0-100 μM). The graph is linear and provides values of $-K_m$ (slope), V_{max} (extrapolation to the y ordinate) and the efficiency constant, V_{max}/K_m, (by extrapolation to the x ordinate). The apparent V_{max}, K_m and V_{max}/K_m values are 50.7 ± 9.5 μM, 1317 ± 159.2 nmol min^{-1} mg^{-1} and 25.9 nmol min^{-1} mg^{-1} μM^{-1}.

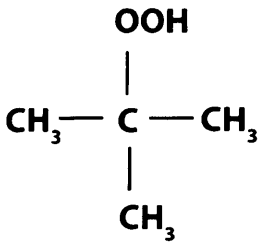
investigation utilising the glutathione-coupled assay.

An interesting point to note is when a cross comparison is made for the efficiency constants V_{max}/K_m of the published data which investigates the substrate cumene hydroperoxide and the value obtained in the present study utilising t-butyl hydroperoxide substrate (V_{max}/K_m of 118 for cumene hydroperoxide; V_{max}/K_m of 25.9 nmol min^{-1} mg^{-1} μM^{-1} for t-butyl hydroperoxide). The substantial difference in values of the efficiency constant for both substrates may be rationalised when the structures of these substrates are considered (Figure 4.22). The most significant difference between the two substrates is that cumene hydroperoxide is aromatic in structure whereas t-butyl hydroperoxide is aliphatic. This may be correlated to the presence of a possible hydrophobic pocket in the TcGPXI tertiary structure where the catalytic tryptophan residue is located. This would suggest that the indole ring of the tryptophan residue facilitates better π π (pi-pi) stacking associated with the ring structure of the substrate cumene hydroperoxide, thus allowing tighter binding compared to the t-butyl hydroperoxide substrate. The catalytic tryptophan has previously been suggested as having an essential role in the orientation of the substrate molecule in some enzymes such as in the cysteine protease Papain enzyme (Hussain et al., 2002) and more relevantly in the catalytic mechanism of the TcGPXI homologue *T. brucei* PxIII (Schlecker et al., 2007).

Similarly to the glutathione-linked enzyme reactions, TcGPXI cysteine mutant enzymes, C48G and C96G, were found to exhibit no catalytic activity, whereas mutation of Cys77 to serine or alanine has been found to allow catalytic activity to be retained. The lack of catalytic activity in the TcGPXI mutant enzymes (C48G and C96G) may suggest that reactive thiol groups of both Cys48 and Cys96 are required for efficient catalytic activity.



CUMENE HYDROPEROXIDE



t-BUTYL HYDROPEROXIDE

Figure 4.22: Chemical structures of cumene hydroperoxide and t-butyl hydroperoxide
The chemical structures of cumene hydroperoxide ($C_6H_5C(CH_3)_2OOH$) and t-butyl hydroperoxide (C_4H_9OOH) substrates for wild type TcGPXI are shown.

4.22 Summary

The results of both the tryparedoxin-linked and glutathione-linked assays, demonstrated as expected, that the wild type TcGPXI enzyme exhibits catalytic activity comparable to that of the published values, whereas the TcGPXI mutant C48G, C77G and C96G enzymes are rendered inactive.

The lack of catalytic ability of C48G and C96G TcGPXI mutant enzymes, may suggest that both Cys48 and Cys96, play a pivitol role for catalytic activity of the TcGPXI enzyme. The lack of activity exhibited by the C77G mutant TcGPXI enzyme may be suggestive of a structural role, which is supported by the 1D ^{1}H NMR spectrum of the protein. If both free thiol groups contributed by Cys48 and Cys96 are required for catalytic activity, this would suggest the possibility of an atypical 2-Cys peroxiredoxin mechanism for wild type TcGPXI (Figure 4.23), similar to that observed in the *T. brucei* homologue PxIII. The possibility of an atypical 2-Cys peroxiredoxin mechanism may be supported by results of triple resonance NMR experiments predicting the presence of a disulphide bond between Cys48 and Cys96 occurring in the oxidised form. *T. cruzi* GPXI and *T. brucei* PxIII show 75% sequence identity where TbPxIII catalytic residues Cys47 and Cys95 are analogous to TcGPXI residues Cys48 and Cys96; thus it is reasonable to postulate that these enzymes share similar structural and mechanistic features.

Initially the TcGPXI C77G mutant enzyme was found to aggregate and exhibit no catalytic activity. Mutation of the TcGPXI Cys77 residue to an alternative residue by site directed mutagenesis, may facilitate the production of a more stable folded protein and allow further NMR studies augmented by detailed kinetic investigation to discover the exact role of Cys77 in TcGPXI. If the mutation of Cys77 to an alternative residue generates a stable folded protein and additionally generates catalytic activity, this may contribute to the confirmation of an atypical 2-Cys peroxiredoxin type mechanism in wild type TcGPXI. Following this work, subsequent investigation of the catalytic activity of TcGPXI where Cys77 has been mutated to alanine or serine (Alaba 2008; unpublished work carried out in our laboratory) has revealed that catalytic activity is not eliminated, suggesting that Cys77 is not essential for catalysis.

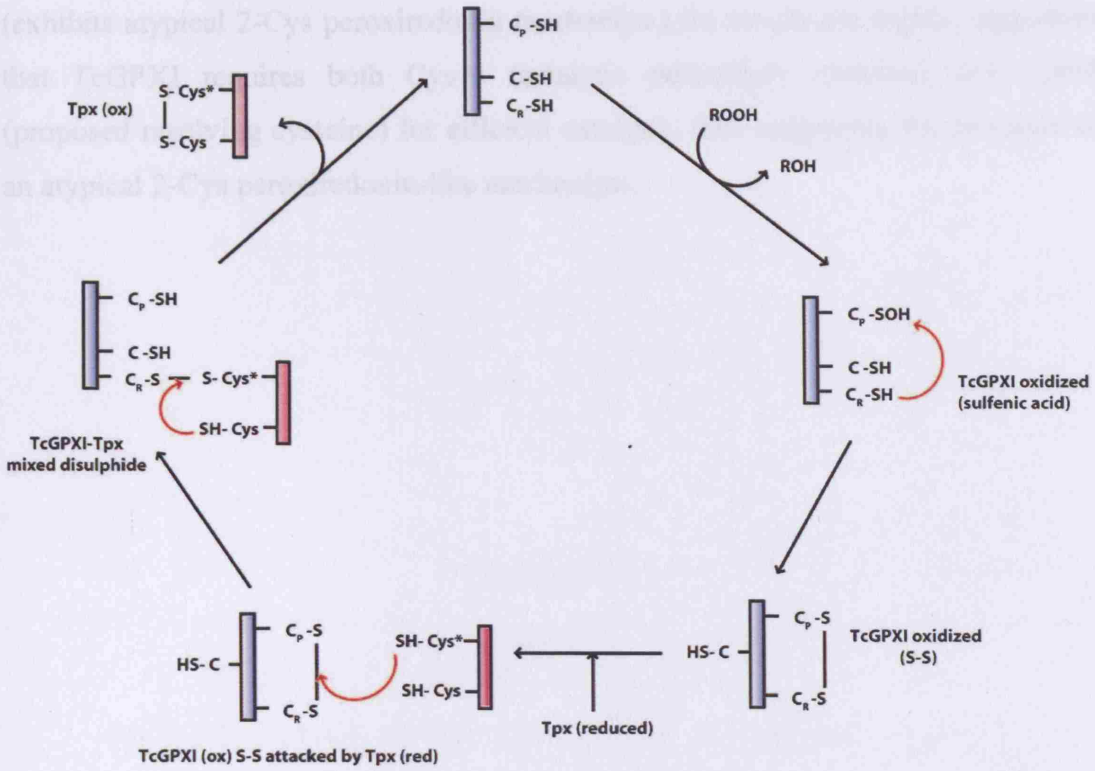


Figure 4.23: Proposed atypical 2-Cys peroxiredoxin mechanism for wild type TcGPXI

Diagrammatic representation of proposed mechanism for wild type *T. cruzi* GPXI; where C_P (peroxidatic cysteine) and C_R (resolving cysteine) represent Cys48 and Cys96 respectively. The first stage is the reduction of the hydroperoxide by the catalytic peroxidatic cysteine (C_P; Cys48), after which nucleophillic attack by the resolving cysteine (C_R; Cys96) results in a disulphide bond (stable oxidised intermediate form). The redox molecule tryparedoxin (Tpx) contains redox-active dithiol group and forms a mixed disulphide, ultimately leading to the regeneration of the reduced form of TcGPXI. The oxidised disulphide form of tryparedoxin may then be reduced by dihydrotrypanothione. (Adapted from Schlecker et al., 2007).

A detailed mutational and kinetic analysis in conjunction with NMR investigation, would be required to extensively research the functional roles of various residues and their effect on catalysis, thus confirming the exact catalytic mechanism of the wild type TcGPXI enzyme. However from the results of this study and with consideration of the high sequence homology to the *T. brucei* PxIII enzyme (exhibits atypical 2-Cys peroxiredoxin mechanism) the results are highly suggestive that TcGPXI requires both Cys48 (catalytic peroxidatic cysteine) and Cys96 (proposed resolving cysteine) for efficient catalysis, thus suggesting the presence of an atypical 2-Cys peroxiredoxin-like mechanism.

Chapter Five

Results

THE CRYSTAL STRUCTURE OF OXIDISED TcGPXI

5.1 Crystallography

This chapter provides a description of the crystallisation of proteins, data collection and the application of molecular replacement in the analysis of X-ray diffraction data of oxidised wild type TcGPXI. Crystallography is vast area of research and mathematical theory, therefore the mathematics will not be focused on but rather the basic application of techniques of going from crystal to structure.

5.2 Growth of protein crystals

Protein crystals are grown by slow, controlled precipitation from aqueous solution under conditions that do not denature the protein. Ionic compounds such as salts and polymer polyethylene glycol are widely used. Slow precipitation is achieved when the precipitant is added to the aqueous protein solution until the concentration of the denaturant is just below that required to precipitate the protein. Evaporation of water slowly occurs in a controlled manner, thereby increasing the concentrations of protein and precipitant, until precipitation occurs. The formation of crystals depends on numerous factors including the protein construct itself, the concentration of the protein, ionic strength, pH and temperature. The crystallisation phenomenon can be illustrated by the phase diagram (Chayen, 2004) which is experimentally obtained by varying two parameters at a time and thus represents a two-dimensional "slice" of the multidimensional space of parameters relevant to crystallisation. The phase diagram (Figure 5.1) shows which state (liquid, crystalline or amorphous solid) is stable under various crystallisation factors and regions corresponding to high supersaturation (where the protein precipitates), moderate supersaturation (spontaneous nucleation), metastable zone (where crystals are stable and may grow but there is no further nucleation) and undersaturation (the protein is fully dissolved and crystals will never grow).

Crystallisation occurs in two phases consisting of nucleation and growth, where the former is an essential requirement for the growth of large crystals suitable for X-ray diffraction. In cases of excess nucleation seeding methods can be used, where nuclei can be directly inserted into the metastable zone for crystal growth.

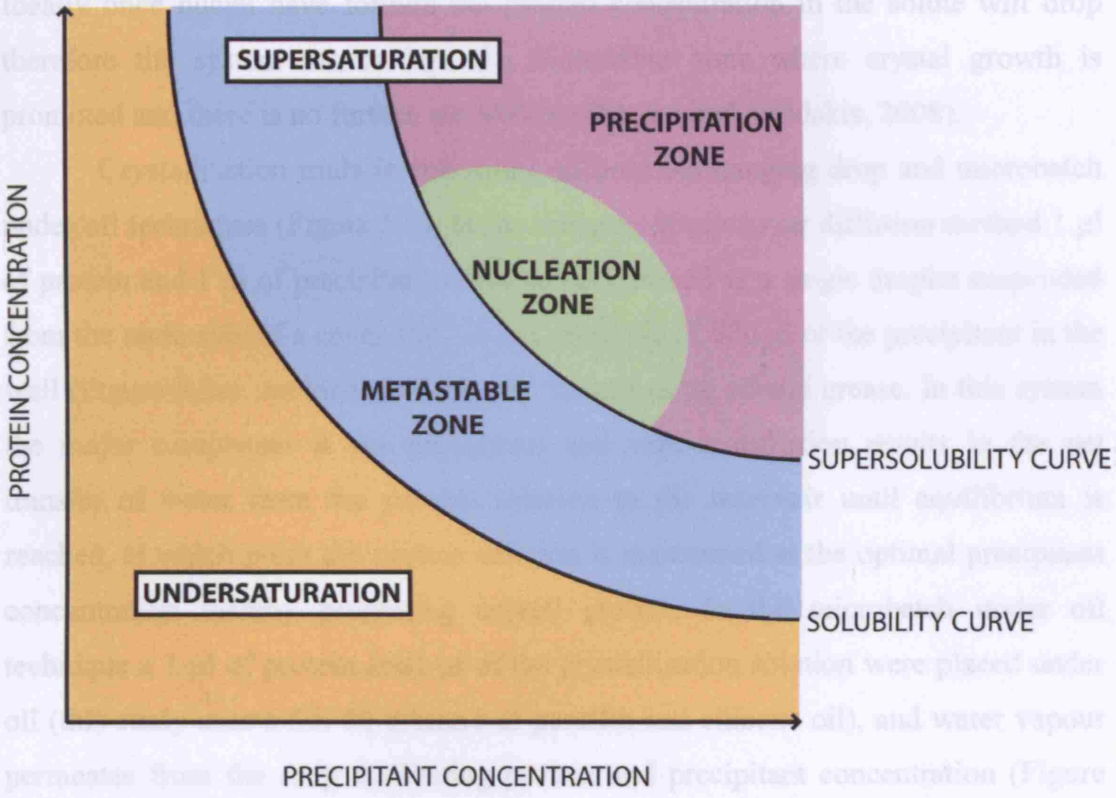


Figure 5.1: Protein crystallisation phase diagram

The diagram illustrates the phase diagram for crystallisation where protein concentration is varied against precipitant concentration. The orange region indicates concentrations where the solution is not saturated with protein thus neither nucleation or growth occurs. The metastable zone (blue), precipitation zone (purple) and nucleation zone (green) indicate unstable regions that are supersaturated with protein. The nucleation zone indicates conditions of nucleation and growth whereas the metastable zone represents conditions where only growth can occur. The solubility curve is defined as the protein concentration in solution that is in equilibrium with crystals; the supersolubility curve is defined as the separation of conditions where spontaneous nucleation occurs from conditions where the crystallisation solution stays clear when undisturbed. (Adapted from Chayen, N. 2004).

In vapour-diffusion methods (e.g. hanging drop methods) the protein solution begins in the undersaturated zone, and reaches supersaturation by equilibration with the solution containing crystallisation reagents (or precipitant). Ideally once nuclei have formed the protein concentration in the solute will drop therefore the system moves into the metastable zone where crystal growth is promoted and there is no further nucleation (Chayen and Saridakis, 2008).

Crystallisation trials in this study utilised the hanging drop and microbatch under oil techniques (Figure 5.2). In the hanging drop vapour diffusion method 1 µl of protein and 1 µl of precipitant solution were placed as a single droplet suspended from the underside of a cover slip, over a reservoir of 500 µl of the precipitant in the well (Figure 5.2a). An air-tight seal was formed using silicon grease. In this system the major component is the precipitant and vapour diffusion results in the net transfer of water from the protein solution to the reservoir until equilibrium is reached, at which point the protein solution is maintained at the optimal precipitant concentration thereby promoting crystal growth. In the microbatch under oil technique a 1 µl of protein and1 µl of the crystallisation solution were placed under oil (this study uses a 50: 50 mixture of paraffin and silicone oil), and water vapour permeates from the drop facilitating protein and precipitant concentration (Figure 5.2b).

5.3 Basic crystallography theory

A crystal is a three-dimensional array of identically oriented molecules held together by noncovalent interactions. The array can be divided into identical unit cells, which are the smallest and simplest volume element that is completely representative of the whole crystal. The corners of the unit cells can be observed as an array of points, where the regular spacing of the origins of the unit cells giving rise to the crystal lattice. Crystallography provides an image of the electron clouds that surround the molecule in the average unit cell of the crystal, in which the location of a specific atom is given by a set of 3D Cartesian coordinates x, y and z. A lattice point that is designated as the origin of the unit cell's coordinate system is thus assigned to have coordinates of 0,0,0.

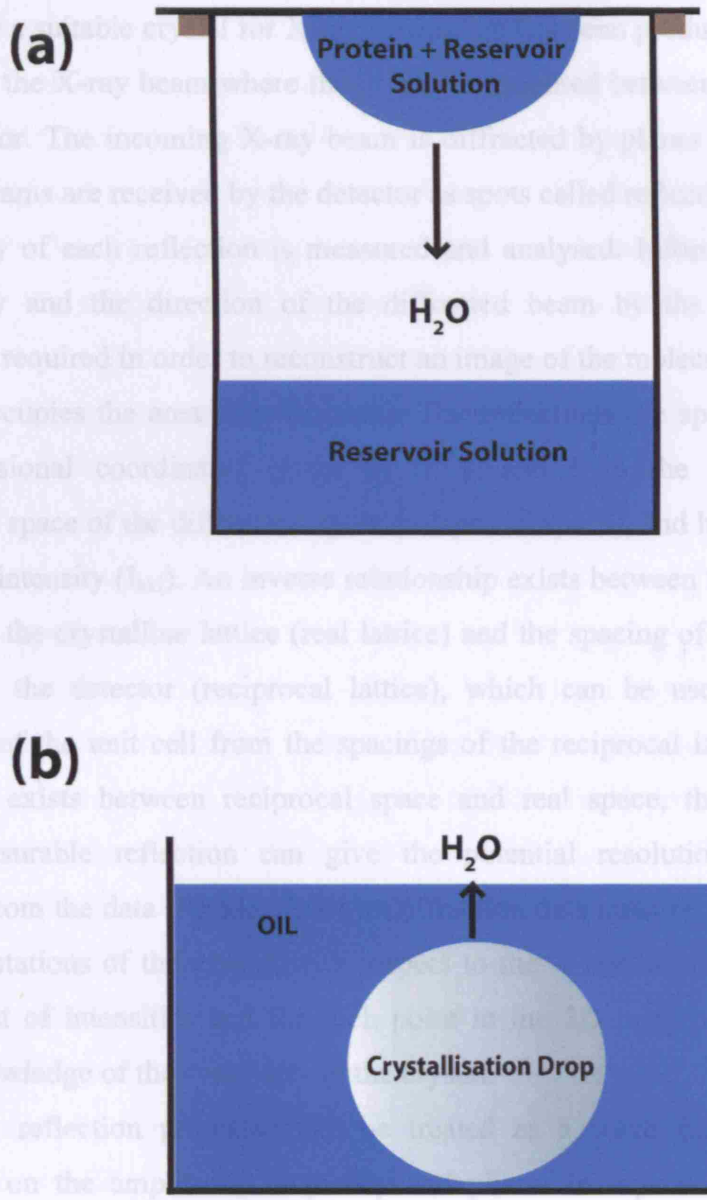


Figure 5.2: Crystallisation techniques

The two crystallisation methods employed to produce protein crystals in this study: (a) vapour-diffusion hanging drop method and (b) microbatch under-oil techniques.

In the hanging drop method 1 µl of protein mixed with 1 µl of precipitant is suspended from a cover slip over a 500 µl reservoir of precipitant, prior to sealing the system with silicon grease. For microbatch under-oil techniques 1µl of protein mixed with 1µl of precipitant is placed under oil.

5.3.1 Overview of the aim of crystallography

Once a suitable crystal for X-ray diffraction has been produced, the crystal is subjected to the X-ray beam where the crystal is mounted between an X-ray source and a detector. The incoming X-ray beam is diffracted by planes of atoms and the diffracted beams are received by the detector as spots called reflections. The position and intensity of each reflection is measured and analysed. Information containing the intensity and the direction of the diffracted beam by the crystal for each reflection is required in order to reconstruct an image of the molecule in the unit cell (unit cell occupies the area of real space). The reflections are specified by sets of three-dimensional coordinates given by h, k and l in the imaginary three-dimensional space of the diffraction spots (reciprocal space), and have an associated measurable intensity (I_{hkl}). An inverse relationship exists between the spacing of the unit cells in the crystalline lattice (real lattice) and the spacing of reflections in the lattice from the detector (reciprocal lattice), which can be used to deduce the dimensions of the unit cell from the spacings of the reciprocal lattice. An inverse relationship exists between reciprocal space and real space, therefore the most distant measurable reflection can give the potential resolution of the model obtainable from the data (Rhodes, 2006). Diffraction data must be collected from all unique orientations of the crystal with respect to the X-ray beam in order to get a complete list of intensities and for each point in the 3D reciprocal lattice, which requires knowledge of the symmetry in the crystal.

Each reflection produced can be treated as a wave function containing information on the amplitude, frequency and phase. In crystallography structure factors provide the wave description of X-ray reflections where each diffracted X-ray that produces a reflection at the detector may be described as the sum of all the contributions of all scatterers (each atom is a scatterer) in the unit cell. For each reflection of indices hkl there is associated structure factor (F_{hkl}), which is the sum of many wave equations (one for diffraction by each atom). This suggests that each reflection contains contributions from diffraction by electron clouds of every atom in the unit cell.

Diffraction data reveals the electron density which can be described by a periodic function $\rho(x,y,z)$ with a value of ρ for electron density at every position (given by coordinates x,y,z) in the unit cell. A graph of the function produces the

image of the electron density map, therefore the ultimate aim is the deduction of the function whose graph is the electron density map (Rhodes, 2006).

The Fourier transform describes the mathematical relationship between an object and its diffraction pattern. The function $\rho(x,y,z)$ can be seen as a Fourier transform of the structure factors, where structure factors contain information on the amplitude, frequency and phase of each reflection. Therefore the intensity of one reflection provides the amplitude of one Fourier term in the series that describes $\rho(x,y,z)$ and the position (hkl) specifies the frequency, however the phase is unknown. Once the electron density function is obtained and graphed, a model must be built that fits the map using prior knowledge of primary sequence and that is structurally realistic.

5.3.2 The unit cell

The unit cell of the crystal is the smallest repeating unit that can generate the crystal with only translation operations, and has the dimensions a, b and c with angles α, β and γ (Figure 5.3a). The unit cells are three-dimensionally stacked where the arrays of points of the unit cell are termed the lattice. The crystal lattice is important as it allows the interpretation of several crystal properties without consideration of the unit cell. The unit cell is bound by planes (100), (010) and (001). Additional planes of regular spacing can be constructed through lattice points and it is these planes that are the sources of diffraction. The planes have designated lattice indices or Miller indices (hkl) which specify sets of planes such that h is the number of planes in the set per unit cell in the x direction, k in the y direction and l in the z direction. These planes will thus divide the unit cell. For example, the planes given by Miller indices (2,3,4) divide the unit cell resulting in the unit cell axis a, b and c being intersected into two, three and four parts respectively. Miller indices can also be negative, where the planes tilt in the opposite direction i.e. (234) and $(\overline{234})$. Another example is the set of planes given by Miller indices (210), which intersect the unit cell dividing the unit cell axis a, b and c into two, one and zero parts respectively (Figure 5.3b). The distance between planes of the same set is given by the term, d. According to Bragg's law, reflections from sets of planes are treated as independent diffractors and produce a single reflection.

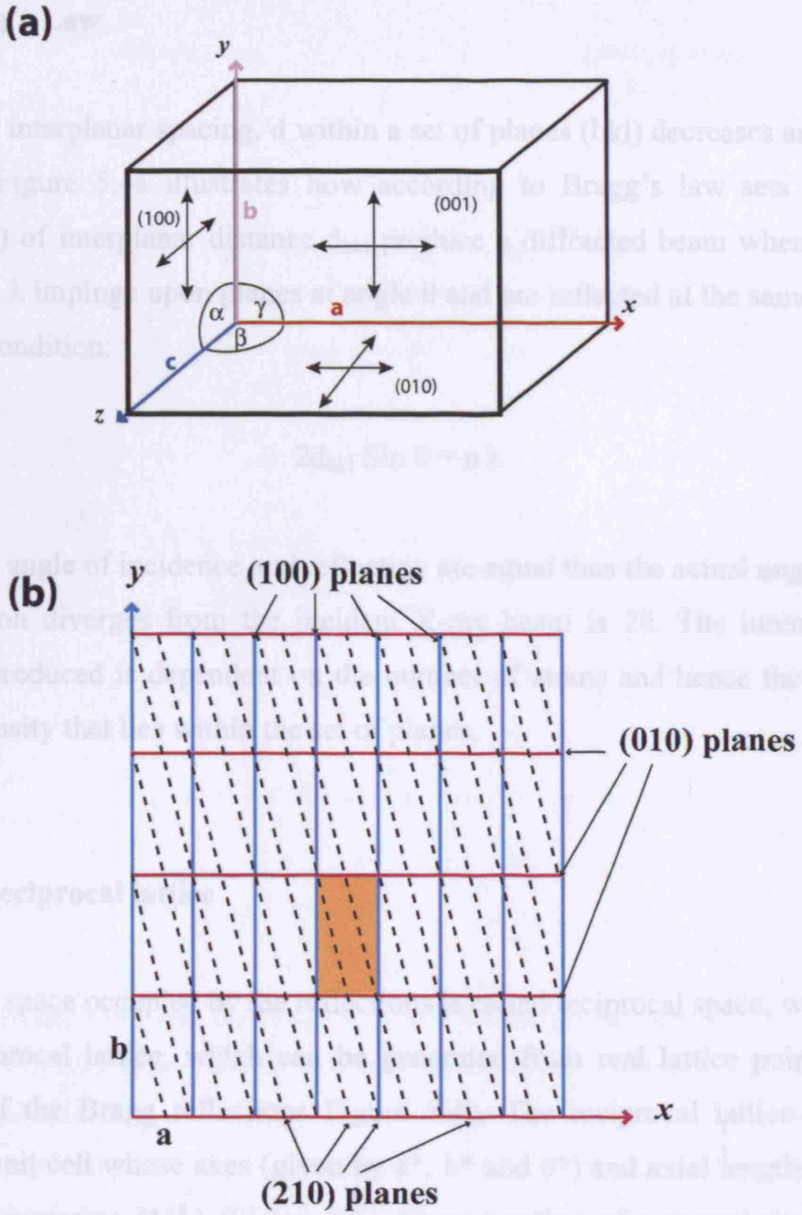


Figure 5.3: General unit cell and atomic planes of a crystal lattice

Unit cell dimensions are defined by lengths **a**, **b**, **c** and the three angles α, β and γ, where the sets of atomic planes in the crystal lattice that are most apparent are those determined by the faces of the unit cell (a). The set of planes parallel and including the **bc** face are given as (100); planes parallel and including the **ac** face are designated (010); planes parallel and including the **ab** face are designated (001). Additional atomic planes (210) in a two dimensional section of the crystal lattice are illustrated (b), where one unit cell is highlighted (orange) and unit cell edges **a** and **b** are shown. The set of planes denoted by Miller indices (210) are shown to intersect the unit cell edge **a** twice, unit cell edge **b** once and not intersect unit cell edge **c**.

5.3.3 Bragg's Law

The interplanar spacing, d within a set of planes (hkl) decreases as any index increases. Figure 5.4a illustrates how according to Bragg's law sets of parallel planes (hkl) of interplanar distance d_{hkl} produce a diffracted beam when X-rays of wavelength λ impinge upon planes at angle θ and are reflected at the same angle if θ meets the condition:

$$2d_{hkl} \text{ Sin } \theta = n \lambda$$

The angle of incidence and reflection are equal thus the actual angle at which the reflection diverges from the incident X-ray beam is 2θ. The intensity of the reflection produced is dependent on the number of atoms and hence the amount of electron density that lies within the set of planes.

5.3.4 The reciprocal lattice

The space occupied by the reflections is called reciprocal space, where points in the reciprocal lattice, which can be generated from real lattice points, are the locations of the Bragg reflections Figure 5.4b. The reciprocal lattice contains a reciprocal unit cell whose axes (given by a*, b* and c*) and axial lengths will be in reciprocal angstroms (1/Å) (Figure 5.5). Upon rotation of a crystal the reciprocal lattice will also rotate as both are spatially connected. A sphere of reflection can be constructed (Ewald sphere) to describe diffraction with respect to the reciprocal lattice (Figure 5.6). The Ewald sphere has a radius of $1/\lambda$ where the incident beam lies along the diameter, and the origin of the reciprocal lattice is set to the point where the incident beam emerges from the sphere. When the crystal is rotated so that a reciprocal lattice point comes into contact with the Ewald sphere, Bragg's law is satisfied and a reflected beam is measured. In a typical diffraction experiment the crystal is rotated in the X-ray beam so that various reciprocal lattice points come into contact with the Ewald sphere, each producing a beam in the direction of a line from the centre of the sphere through the reciprocal lattice point, and each reflection produced is from a set of real-space planes. Therefore, the number of reflections and

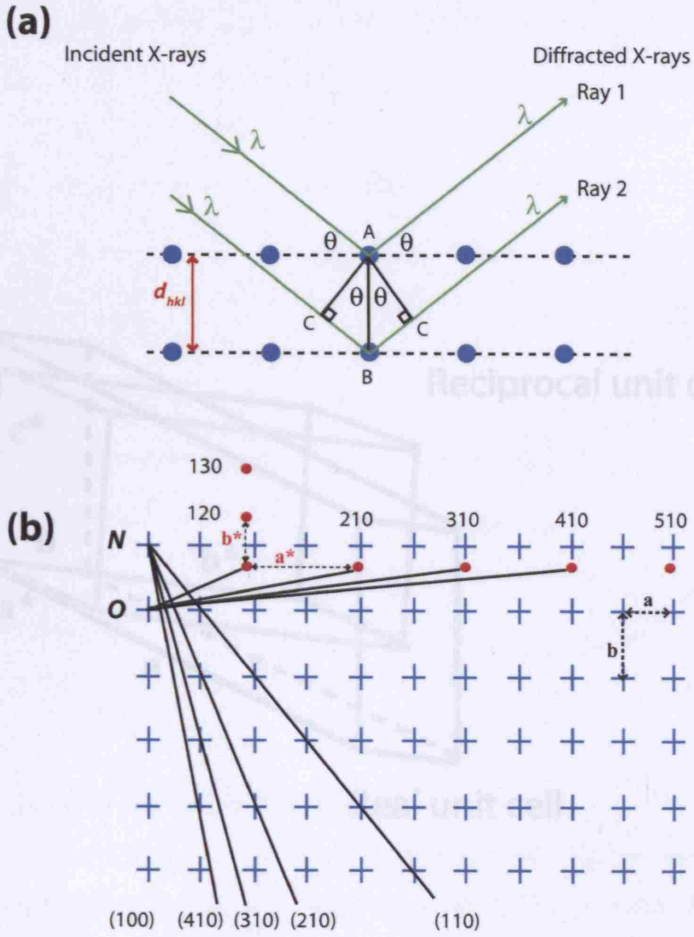


Figure 5.4: Bragg's law and the construction of the reciprocal lattice

(a) The conditions that produce diffraction of the X-ray beam from equivalent parallel planes according to Bragg's law, where blue dots represent two parallel planes of lattice points. Sets of parallel planes with index *hkl* and interplanar distance d_{hkl} produce a diffracted beam when X-rays of wavelength λ impose at an angle θ and reflect with the same angle, if θ meets the condition $2d_{hkl} \sin\theta = n\lambda$, where n is an integer. Satisfaction of Bragg's law allows reflection of Ray 1 and 2 at angle θ. If Ray 2 travels an additional distance that is an integral multiple of λ, then Rays 1 and 2 follow constructive interference.

(b) Construction of the reciprocal lattice, where crystal lattice points are shown as (+) and reciprocal lattice points are shown as (•). Lattice planes (110), (210), (310), and (410) are shown as lines emerging from point N. The reciprocal lattice points are generated from construction of lines that are drawn normal to these planes from origin O at a distance of $1/d_{hkl}$ from O, where d_{hkl} is the perpendicular spacing between lattice planes of indices hkl. Real unit cell edges are given as **a** and **b** in addition to reciprocal unit cell edges **a*** and **b***.

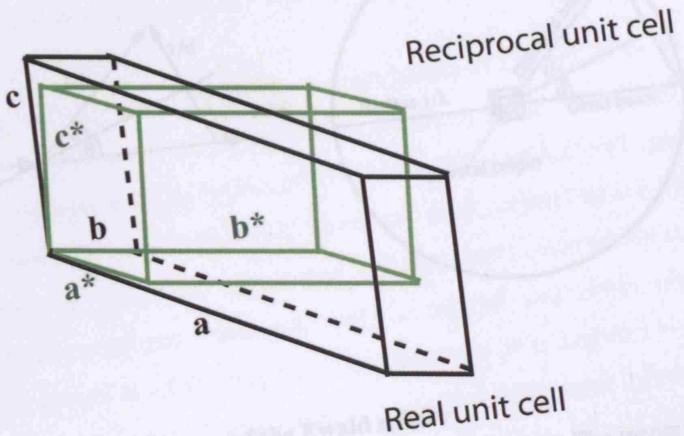


Figure 5.5: Real and reciprocal unit cells

Example of real (black) and reciprocal (green) unit cells in the monoclinic system where in the monoclinic system $\mathbf{a} \neq \mathbf{b} \neq \mathbf{c}$, $\alpha = \gamma = 90\,°$, $\beta \neq 90\,°$. Reciprocal cell edges are given as **a***, **b*** and **c***, whereas real unit cell edges are given as **a**, **b** and **c**. Reciprocal cell edge **a*** is perpendicular to real space plane **bc**, **b*** is perpendicular to real space plane **ac**, **c*** is perpendicular to real space plane **ab**.

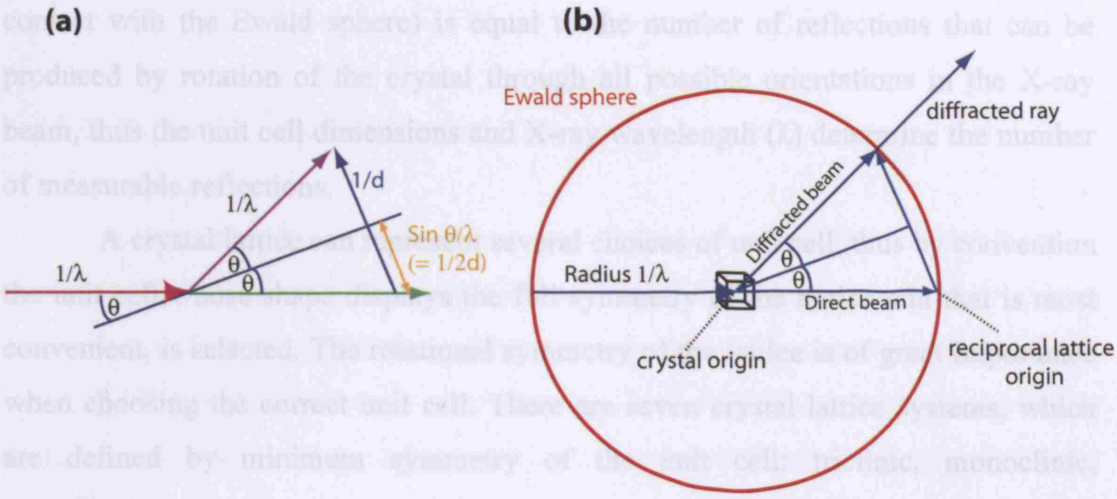


Figure 5.6: The construction of the Ewald sphere

A simplistic view of the diffraction that occurs is shown in (a). The incoming ray (pink) and diffracted ray (purple) are at an angle of θ due to a set of Bragg planes in the crystal. The reciprocal space vector (dark blue) is perpendicular to the Bragg planes and is given by the difference between the undeflected beam (green) and diffracted ray (purple), with a length 1/d. This can be extended to any set of Bragg planes where the reciprocal space vector will always span from the direct beam to the tip of the vector at the diffracted ray position. The angle 2θ can be 0-180 ° where the diffracted ray can occur in any direction in three-dimensional space, thus the tip of the reciprocal space vector lies anywhere on the surface of a sphere with radius 1/λ i.e. Ewald sphere (b). As the crystal is rotated in the X-ray beam, reciprocal lattice points will come into contact with the Ewald sphere, generating a diffracted beam from the centre of the sphere through the reciprocal lattice point that is in contact with the Ewald sphere. For example, an hkl reflection is produced when reciprocal lattice point X_{hkl} comes into contact with the Ewald sphere.

their direction is dependent on the dimensions of the unit cell and not the unit cell contents (Rhodes, 2006). The number of reciprocal lattice points within the limiting sphere (any reciprocal point within distance of 2/ λ of origin can be rotated into contact with the Ewald sphere) is equal to the number of reflections that can be produced by rotation of the crystal through all possible orientations in the X-ray beam, thus the unit cell dimensions and X-ray wavelength (λ) determine the number of measurable reflections.

A crystal lattice can represent several choices of unit cell, thus by convention the unit cell whose shape displays the full symmetry of the lattice and that is most convenient, is selected. The rotational symmetry of the lattice is of great importance when choosing the correct unit cell. There are seven crystal lattice systems, which are defined by minimum symmetry of the unit cell: triclinic, monoclinic, orthorhombic, tetragonal, rhombohedral, hexaganal and cubic. For example a monoclinic system is where unit cell dimensions $a \neq b \neq c$, angles $\alpha = \gamma$ but angle $\beta \neq 90°$. Normally a unit cell is represented by a single lattice point located at a corner of the unit cell; however points may be added at the centres of one or all three unique unit cell faces (face-centring) or at the centre of the unit cell (body-centring). This results in the further classification of the 14 Bravais crystal lattices. The dimensions of the unit cell can be inferred from the distancing between reflection spots from the diffraction experiment.

The symmetry in the unit cell will mean that the lattice will also be symmetrical, thus sets of reflections are equivalent. Knowledge of the unit cell symmetry can reduce the amount of data required to be collected and facilitates the formulation of a strategy for data collection. The internal symmetry of the unit cell is described by its space group, the nomenclature of which contains a capital letter (representing the lattice type) and other symbols (representing symmetry operations). There are 230 space groups and 14 allowable Bravais lattices. Once the lattice type is described the internal symmetry of the unit cell is required. The simplest symmetry operators are translation, rotation and reflection which can be combined to describe complex symmetry elements including screw axis and glide planes. Due to protein chirality (mirror planes are incompatible as natural proteins are only L-amino acids) protein crystals include only translations, rotations and screw axis (rotation and translation combined), which reduces the number of possible space groups from 230 to 65. The unit cell contains identical molecules

arranged in symmetry of which the largest molecule containing no symmetry is called the asymmetric unit. A regular absence of reflections resulting from symmetry facilitates the determination of the space group (Rhodes, 2006).

5.3.5 Collection of X-ray diffraction data

In the diffraction experiment a monochromatic incident X-ray beam from an X-ray source is diffracted by the crystal into distinct beams. These are received by a detector as reflections (diffraction spots). Elements such as copper provide a good source of X-rays however most commonly and well utilised are those sources from large particle storage rings such as synchrotron sources. These large ring structures contain electrons or positrons that are accelerated near to the speed of light in circular motions by powerful magnets. The energy emitted is in the form of X-rays, and the beam is directed through a collimator and subjected to focusing mirrors in order to produce monochromatic light.

The crystal is mounted onto a goniometer (after bathing in suitable cryoprotectant) that allows precise changes in the crystal position in order to bring any reciprocal lattice point(s) into contact with the sphere of reflection, and thus diffracted rays are produced and received by the detector which is able to record the intensity of the diffracted beam. Modern high-speed data collection relies on a rotation/oscillation method whereby the crystal is rotated and oscillated back and forth. Data collection software is used to determine the crystal orientation from frames over a small rotation range. The reflections are indexed, and unit cell dimensions and space group are determined in order to establish the best strategy for data collection. Data collection is carried out to obtain a set of measured, indexed intensities for as many reflections as possible and scaled so that the dataset contains identical intensities for identical reflections. The strategy for data collection allows crystal orientations to be used for data collection that contain unique reflections, from deduction of the unit cell space group determined from regular absences in the diffraction pattern (half of the reflections in the reciprocal lattice will be redundant). Therefore, data collection over 180° will provide unique reflections. The angle of data collection can further be reduced by consideration of additional symmetry elements *i.e.* the addition of a centre of symmetry reduces the possible 230 space

groups to 11 different groups termed the Laue groups. For each Laue group it is possible to calculate the number of reflections that are unique from data collected by a particular rotation of the crystal in the X-ray beam. The redundancy of data is used to improve the signal to noise ratio by averaging the multiples of equivalent reflections.

5.3.6 The phase problem

A measured reflection can be described by a structure factor equation, containing one term for each atom in the unit cell. The electron density map can be described as the Fourier sum in which each term is a structure factor, thus a Fourier transform of complete structure factors provides the function ($\rho(x,y,z)$) whose graph results in the electron density map. A structure factor is a periodic function whose amplitude is proportional to the square root of the intensity of a reflection produced from a set of planes ($(I_{hkl})^{1/2}$). The frequencies of the three dimensional wave function are h, k and l (indices of the set of planes producing the reflection), thus the frequency of a structure factor is equal to $1/d_{hkl}$, which makes the wavelength the same as the spacing of the planes producing the reflection (Rhodes, 2006). However phase information for each reflection is lost and is required to compute the complete structure factors in order to generate the electron density map into which a model can be built. Techniques such as isomorphous replacement, anomalous scattering and molecular replacement are commonly used to overcome the phase problem. These techniques provide initial phases that must be improved through refinement in order to produce the interpretable electron density map.

5.3.7 Molecular replacement

The technique of molecular replacement uses initial phases from the structure factors of a known homologous protein. A good model must be chosen *i.e.* one that models a high proportion of the scattering from the target structure with high accuracy and that has a high sequence identity. The model structure template

can be improved by the removal of sequence regions exhibiting high diversity (*e.g.* loop regions) and truncating different side chains to common atoms (*e.g.* alanine).

Initial phases are obtained by placement of the known structure model in the unit cell of the unknown protein to create the best phasing model. The best orientation of the model protein can be found using Patterson-map comparisons and then using structure factor comparisons to find the best location of the model protein. The original concept of the molecular replacement method (Rossmann and Blow, 1962) utilises three stages (Rossmann, 1990):

1. Determination of the angular orientation of the search molecule where orientation of the search molecule produces maximal overlap with the target structure in the absence of phase information (rotation search);
2. Determination of the position of the search molecule in 3D space relative to symmetry elements of the target crystal (translation search);
3. Phase determination using knowledge of symmetry operators.

Prior to molecular replacement it is useful to estimate the number of molecules in the unit cell and hence the number of molecular replacement solutions. Matthews determined the solvent content of 116 protein crystals to be typically 50% (Matthews BW., 1968). Using Matthews coefficient (V_m), the volume of the unit cell and molecular weight of the protein it is possible to ascertain the number of molecules in the asymmetric unit and thus the number of molecules in the unit cell.

The Patterson function is a Fourier transform of the squared structure factor amplitudes with phases set to zero, and can therefore be calculated for both the model and the observed experimental data. The function produces a contour map that can be calculated with prior knowledge of structure without the inclusion of phase angles. In a molecular replacement rotation search, the model is rotated about three orthogonal axes x, y and z and the resulting Patterson maps compared to the experimental Patterson maps for the target (cross-rotation search). Following the conclusion of the correct orientation, a translation search is carried out to identify its precise position with respect to the origin of the unit cell. This involves translation along the x, y and z axes to find where the structure factor amplitudes of the model are most similar to those determined for the observed data.

The initial phases are subsequently taken from the model to calculate the first electron density maps from the electron density function $\rho(x,y,z)$ which is the Fourier sum of the structure factors. The electron density function contains information of the amplitudes of each hkl ($|F_{obs}|$), which is proportional to the square root of the intensity recorded and weighted phases are calculated. The model can be changed to improve the phases and as the maps become clearer so the model becomes more detailed. Phases determined by molecular replacement methods can suffer model bias therefore nF_o- F_c maps can be calculated and visualised using computer graphics programs such as COOT (Emsley and Cowtan, 2004). This protocol subtracts calculated structure factor amplitudes from a multiple of the observed structure factor amplitudes within each Fourier term. The F_o- F_c map contains positive and negative areas in the electron density map, where positive regions represent where the contribution of observed intensities are greater than the model thus the model must be adjusted to increase the density in this area by movement of atoms. The negative areas in the F_o- F_c map indicates the presence of more electron density in the unit cell than it actually contains therefore atoms need to be moved out of the negative density. In this way nF_o- F_c maps can help pinpoint regions of error. Another technique is the use of omit maps, where atoms in a questionable region are eliminated and structure factors calculated so that possible errors in the region contribute nothing to the phase, thus not biasing the resulting map.

5.3.8 Structural refinement

Refinement is the cycle of map calculations and model building where forms of real space refinement of the model are interspersed with computerized steps to improve the agreement of the model with the original experimental intensity data. This process involves comparison of computed with observed structure factor amplitudes and therefore termed reciprocal space refinement. Ultimately the goal is to minimise the differences between F_{obs} (observed structure factors) and F_{calc} (calculated structure factors).

The initial phases derived from molecular replacement using a known structure can often lead to model bias, where the underlying cause is due to the low

data to parameter ratio resulting from a lack of resolution in the dataset. Through the process of refinement overfitting to the diffraction data lead to the introduction of systematic errors. The observed data to parameter ratio can be improved by using constraints (reduction in the number of parameters) or restraints (increasing the number of observations) of bond lengths and bond angles, bearing in mind the geometry of the model, which is particularly important in the refinement of low resolution datasets.

The earliest refinement methods were based on least-squares methods. Currently, there are two types of refinement commonly used: Maximum Likelihood methods (implemented in the REFMAC (Murshudov et al., 1997) refinement program of the CCP4 software suite) and Molecular dynamics (Simulated Annealing technique applied in CNS (Adams et al., 1997)).

5.3.8.1 Least squares

The earliest successful refinement technique was a version of the least-squares method. The approach of least squares methods is to minimise the function:

$$\Phi = \sum_{hkl} w_{hkl} (F_o - F_c)^2_{hkl}.$$

where w_{hkl} is a weighting term; F_o is observed structure factor amplitude for reflection hkl; F_c is calculated amplitude for the structural model for the same hkl reflection. Thus the function Φ is the sum of the squares of differences between observed and calculated amplitudes. The sum is taken over all the hkl reflections and weighted.

Various parameters are used in refinement and required to minimise the function Φ *e.g.* atomic positions. Other parameters are occupancy (usually set to 1.0), which is a measure of the fraction of molecules in which an atom occupies a specific position in a model. The temperature factor or B-factor for each scatterer (atom) or groups of scatterers describes the degree to which the electron density is spread out showing the dynamic mobility of atoms, and can also be refined. B-factors are also indicative of structural errors in the model, *e.g.* if an area of model is

incorrectly built then B-factors will be higher than those of correctly built atoms nearby. The refinement of B-factors is dependent on the resolution of the data as well.

From the complicated function Φ, there are many local minima states that relate to variations in model conformation that minimise the function Φ with respect to other neighbouring similar conformations. The method of least squares finds the minimum closest to the starting point thus model parameters that are close to the global minimum (conformation of best agreement with native structure factors) are desirable. Approaches to increasing the probability of finding the global minimum include the application of constraints and restraints of parameters, with penalization if the model becomes unrealistic.

5.3.8.2 Maximum likelihood

The minimisation of least squares function can lead to systematic errors in the model by fitting noise in the diffraction data and thus the least squares residual is justified only if the model is nearly complete and error-free. In the initial stages of refinement improved refinement techniques for incomplete and error-containing models utilise the maximum likelihood formulation. Modern refinement programs such as the program REFMAC in the software suite of CCP4 uses the maximum likelihood techniques. The principle uses the concept that the quality of the model is judged according to its coherence with an observation, thus the probability will be high for regions where the model is built correctly. For independent observations, the joint probability of making a set of observations is the product of the probabilities of making each independent observation (Murshudov et al., 1997). In crystallography the function represents the probability of obtaining an observed structure factor given a calculated value.

In the maximum likelihood method implemented in the refinement program REFMAC, the function is extended to include prior phase information and experimental standard uncertainties, whilst considering the fact that different structural regions contain different errors. In general, an empirical approach is used to estimate the model errors. Structure factor distributions depend on a single parameter σ_A (the fraction of the normalised structure factor that is correct) where

both the centre of the 2D Gaussian distribution and its width are affected by σ_A therefore the quality of maximum likelihood refinement is sensitive to the quality of the estimation of σ_A i.e. as σ_A decreases so the size of the Gaussian error increases. Model errors have a bigger effect on structure factor agreement at high resolution than at low resolution data and therefore σ_A is a function of resolution. Probabilistic methods for estimating the effects that model errors have on structure factors enable a decrease in model bias by the calculation of σ_A weighted electron density maps. The σ_A distribution for the model is a good source of error estimation, however as the σ_A distribution changes so the error estimates must be updated therefore improved calculated model structure factors will enhance the next refinement cycle. For the maximum likelihood technique to be an improvement on the least squares residual, cross validation has been found to be essential (Murshudov et al., 1997). Development of an efficient Gaussian approximation in the case of structure factor amplitudes without prior phase information is called the "MLF" target function. In the limit of the perfect model the maximum likelihood function decreases traditional least squares residual with $1/\sigma^2_o$ weighting. A method of estimating σ_A using "free" reflections (5 % of the observed data is left out from refinement) is used in the program REFMAC. This method for refinement thus provides an improved method for the estimation of errors by using Bayesian statistics to model more accurately the error distribution thus the partial nature of the model is more accounted for compared to the least-squares technique.

5.3.8.3 Molecular energy and motion in refinement

The problem with use of maximum likelihood and least squares refinements is that refinement can sometimes lead into getting "trapped" into local minima whilst trying to reach the global minimum. The greatest distance from the global minimum from which refinement will converge correctly is called the radius of convergence, therefore increasing the radius of convergence by the addition of restraints and constraints can decrease the probability of falling into local minima (false minima).

Molecular dynamics simulates movement of molecules by solving Newton's law of motion for atoms moving within force fields that represent covalent and non-

covalent bonding. The energy term related to the difference between the measured (observed) and calculated (from the model) reflection intensities is included, where the energy of the model decreases as its fit to the native dataset improves. In the method of simulated annealing (implemented in CNS), the model is "heated" to a high temperature and thus allowed to move out of local minima followed by slow cooling to find the lowest energy conformation whilst agreeing with the diffraction data. Macromolecular structure calculation and refinement is the search for the global minimum of target function (Brunger and Adams, 2002):

$$E = E_{chem} + w_{data}\, E_{data}$$

where E_{chem} represents empirical data about chemical interactions between atoms in the model and is a function of all atomic positions including information on covalent and non-bonded interactions; E_{data} is the difference between the observed and calculated data; w_{data} is the weighting chosen to balance the gradients arising from both terms.

The simulated annealing algorithm requires a method of generating the Boltzmann distribution at a given temperature T, and an annealing schedule of temperatures sequences at which the Boltzmann distribution is generated. The two widely used mechanisms are Metropolis Monte Carlo (movements are made in coordinate space and new energies are calculated, with acceptable movements related to the temperature) and molecular dynamics (simulates atom motions by propagating in time the classical equations of motion).

Molecular dynamics uses integrations of the equations of motion:

$$m_i\left(\frac{\partial^2 r_i}{\partial t^2}\right) = -\nabla_i E$$

where m_i is the mass of atom I; r_i is the coordinates of atom i; δt is the time step; E is the optimization target. The equations are solved using finite difference methods. Acceleration can be used to update the velocity, which can be used to update atomic coordinates if appropriate small timesteps are taken. The current temperature of the

simulation (T_{curr}) is calculated from the kinetic energy of the molecular dynamics simulation:

$$E_{kin} = \sum_{i=1}^{natoms} \frac{1}{2} m_i \left(\frac{\partial r_i}{\partial t}\right)^2$$

$$T_{curr} = \frac{2E_{kin}}{3nk_b}$$

The temperature can be controlled by scaling:

$$\left(\frac{\partial r_i}{\partial t}\right)' = \left(\frac{\partial r_i}{\partial t}\right)\sqrt{\frac{T}{T_{curr}}}$$

The target function depends on several atomic parameters (atomic coordinates, B-values, occupancies). The large number of variable parameters (at least three times that of the number of atoms) gives rise to complicated target function and results in the multiple minima problem. This problem can be alleviated using method of simulated annealing.

Simulated annealing is a process whereby a solid is heated so that particles randomly arrange in the liquid phase, upon which the system can be slowly cooled so that particles adopt a state of the lowest energy conformation. The simulated annealing method increases the probability of finding a more optimal solution than gradient-descent methods as motion against the gradient ("uphill") is allowed, the likelihood of which is determined by temperature. Simulated annealing functions because local conformational changes lower potential energy, thus increasing kinetic energy, which is removed by temperature control. Once in a lower energy position (*i.e.* in electron density) it is unlikely to jump over the local energy barrier back to the wrong position (Figure 5.7).

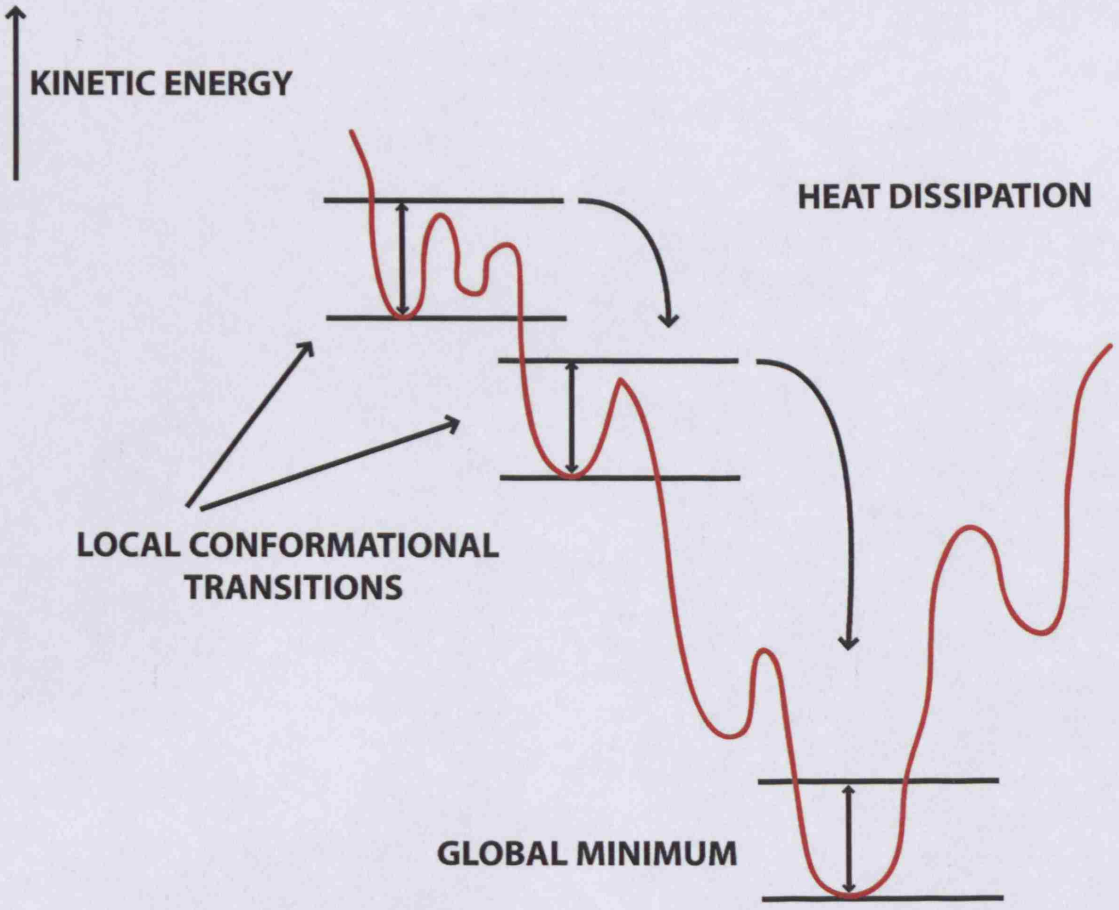


Figure 5.7: Energy diagram for molecular dynamics based simulated annealing
Increasing the kinetic energy of the system allows conformational changes where the energy barriers that trap the system into local transitions are overcome. As the system cools and energy levels drop the system is able to transit and localise at the global minimum. It is unlikely that the system is able to climb out of the global minimum once achieved.
(Based on Brunger and Adams, 2002).

The software suite CNS provides two types of dynamics: Cartesian (chemically restrained) and Torsion angle (chemically constrained) dynamics, of which Torsion angle dynamics is appropriate for most structures. Both types of dynamics provide tight temperature control by velocity scaling (temperature ranges for Cartesian and Torsion angle dynamics are 1000 - 3000K and 1000-10000K respectively). Following the heating up of the system, slow cooling is applied to find the lowest conformational energy state whilst still in agreement with the diffraction data. The simulated annealing protocol can be improved by use of maximum likelihood function targets and cross validation. Torsion angle molecular dynamics simulated annealing with cross validated maximum likelihood targets has been shown to produce significantly less model bias than other methods (Brunger et al., 1998), where this combination has shown to increase the radius of convergence thus allowing productive refinement of poor initial models.

5.3.8.4 Convergence and model quality

The progress of iterative real-space and reciprocal space refinement is monitored by comparison of observed structure factor amplitudes and calculated structure factor amplitudes, where calculation of phases through each stages of refinement reveals what the intensities of the current model would be if it was correct. As the correct structure converges so the observed and calculated structure factors should also converge. The most widely used measurement of convergence and thus an indication into the progress of refinement is the residual index (R factor):

$$R = \frac{\Sigma||F_{obs}| - |F_{calc}||}{\Sigma|F_{obs}|}$$

where F_{obs} is calculated from the observed reflection intensity and each F_{calc} is the amplitude of the correlating structure factor from the model.

An R-factor of greater than 0.5 suggests that the agreement between observed and calculated data is poor and unless more data is provided there will be no improvement. The final desirable R-factor for a protein model will depend on the resolution of the data but typically an R-factor of 0.25 or less, is desirable. A true assessment of the convergence and quality of the model is obtained by the free R-factor (R_{free}). This is computed using a small test set (usually 5 % of the data) which are excluded at the start from refinement. At any stage in refinement the R_{free} measures how well the current model predicts a subset of the measured intensities that were excluded from refinement, whereas R_{factor} measures how well the current model predicts the entire dataset producing the model. The R_{free} is considered to provide the best and least biased assessment of the quality of the model (Adams et al., 1999), however it is routine to assess convergence and model quality based on both the R_{factor} and R_{free}. In addition to these measurements it is also extremely important to monitor the feasibility of the model stereochemically and conformationally, where realistic bond lengths and angles are ensured by use of programs such as the Ramachandran plot. The final model can be validated using programs such as PROCHECK.

5.4 Crystallisation of TcGPXI

Crystals of substrate (t-butyl hydroperoxide) treated oxidised wild type TcGPXI suitable for X-ray diffraction were grown at 20 °C using the hanging drop vapour-diffusion method, where 1 µl of substrate treated protein (at a concentration of 15 mg/ml) and 1 µl of precipitant were mixed suspended as a droplet over a 500 µl precipitant reservoir. Crystals were produced in conditions of unbuffered 1.8 - 2.0 M ammonium sulphate (Figure 5.8). In addition, untreated native wild type TcGPXI crystals suitable for X-ray diffraction were also produced using hanging drop techniques in conditions of unbuffered 1.8 M ammonium sulphate.

In order to prepare a chosen crystal for an X-ray diffraction experiment, the crystal was picked up in a small loop and placed in cryoprotectant of 20% glycerol. The loop containing the crystal in its suitable cryoprotectant was then mounted onto the goniometer and positioned in a stream of cold nitrogen gas for the X-ray experiment.

Figure 5.8: Crystals of oxidised TcGPXI

Crystals suitable for X-ray diffraction were produced in conditions of 1.8 M, 1.9 M and 2.0 M unbuffered ammonium sulphate at 20 °C using the hanging drop vapour-diffusion method.

5.5 Data collection and processing overview

The following section provides a brief overview for the procedures of data collection and processing.

All data was obtained in-house using an R-AXIS IV image plate detector system and Osmic mirrors, in conjunction with the CrystalClear software suite. The initial stages of data collection involve the collection of initial images in order to designate the unit cell and respective dimensions to maximise the strategy for full data collection of all measurable reflections. The first module (module: "Find Spots") in assigning the unit cell initially locates reflections from a single or collection of images, where approximately 50-70 reflections are suggested to be utilised, however this may vary depending on the properties of the image producing the reflections and their location in reciprocal space. The sigma level is specified (sigma level 3.0 was used in this study) in order to account for the average background before a pixel may be considered as a peak. Following the detection of required reflections (also referred to as spots), the procedure of indexing is carried (module: "Index Spots") out whereby the reflections are indexed to elucidate the unit cell dimensions, crystal orientation and Bravais lattice type. The next stage in assigning the unit cell is the refinement of the crystal, detector and source parameters (module: "Refine Cell") thus reducing the difference between observed and calculated reflections, providing validity of the refinement. The size and pixel coordinates of reflections change as they pass through the Ewald sphere of reflection due to crystal mosaicity, source divergence and bandwidth therefore the observed 2D centroids of a Bragg peak will vary over a range of images on which that reflection is found. In the final stage of assigning the unit cell, a reflection list is created containing the calculated reflection centroids. Reflections are predicted (module: "Predict spots") based upon an orientation matrix and the unit cell with corresponding images. Together, these stages provide the assignment of the unit cell and optimal strategy for full data collection.

Following the collection of images of a full dataset from the X-ray experiment, the data is integrated (module: "Integrate Reflections"). Integration of the reflections facilitates the prediction of spots that would appear on a range of images that can then be merged. The mask file is defined which is the shadow of the beam stop and any ice rings may be removed within a defined resolution. The

integration module implemented by d*TREK in the CrystalClear processing software suite processes the collection of images from collected frames and generates a list of intensities describing the X-ray diffraction data collected. The crystal lattice parameters are used to predict the location of the Bragg reflections of the images. The reflection list is indexed and assigned Miller indices (*h,k,l* values). Following integration, the data is subjected to a series of analytical checks (module: "Analyse data") where the Laue group, centricity and space group is assessed and verified.

The final stage of data processing involves the scaling and averaging of the dataset (module: "Scale and Average"). In this module, scale factors are calculated and applied to different batches of reflection from the reflection list. Symmetry equivalent reflections are averaged and the merging and data completeness statistics are calculated. The final output is a list of indexed and unique list of reflections for the crystal.

5.5.1 Data collection for TcGPXI

Data for substrate-treated wild type TcGPXI crystals was collected over an angle of 180° with 1° oscillations at a resolution of 2.3 Å. The data for substrate treated TcGPXI over 82 frames (batches of reflections) was integrated, merged, scaled and indexed using the program d*TREK (Pflugrath, 1999) in the CrystalClear software suite (all X-ray data was kindly collected and processed with the help of Dr. Gary Parkinson at the London School of Pharmacy) producing a final list of intensities for unique reflections. Table 5.1 shows the summary of the data collection statistics for both the native and substrate treated wild type TcGPXI crystals. Data from the substrate treated wild type TcGPXI was carried forward for molecular replacement.

The results of data collection revealed the presence of a C2 space group, with unit cell dimensions: a = 73.846, b = 46.5906, c = 61.1631, $\alpha = \gamma = 90.0$, $\beta = 121.5428$. Calculation using the unit cell volume and Matthews coefficient indicated the presence of one molecule per asymmetric unit.

The file containing the merged data of intensities of indexed unique reflections was imported and the structure factor amplitudes calculated from the

Crystal Parameters	Native TcGPXI	Substrate-treated TcGPXI
Space group	C 1 2 1	C 1 2 1
a (Å)	73.77	73.846
b (Å)	47.02	46.591
c (Å)	61.26	61.163
β	121.78	121.54
No. of independent molecules	1	1
Data Collection Statistics		
Resolution (Å)	23.51 - 3.0 (3.11 - 3.0)	31.47 - 2.3 (2.38 - 2.3)
Wavelength (Å)	1.5	1.5
Temperature (K)	100	100
Rmerge	0.073 (0.229)	0.05 (0.233)
Output <I/σI>	9.5 (3.8)	11.1 (2.7)
Total Reflections	7722	18867
Unique Reflections	3112	7515
Completeness (%)	84.9 (92.5)	94.1 (85.7)
Average redundancy	2.48 (2.39)	2.51 (1.77)
Refinement		
R-factor (%) *	-	20.7 (26.1)
R-free (%) *	-	27.2 (31.6)
r.m.s.d. from ideal geometry:		
Bond lengths (Å)	-	0.021
Bond angles (°)	-	2.031
No. of protein atoms	-	1187
No. of water molecules	-	75
Average B factor (Å^2):	-	31.241

* $R_{factor} = \Sigma|F_o - F_c| / \Sigma F_o$; R_{free} was calculated with 4.7 % of data withheld from refinement

Table 5.1: Data collection and refinement statistics

A summary of the data collection statistics for both native oxidised TcGPXI (untreated with t-butyl hydroperoxide substrate) and substrate treated TcGPXI (t-butyl hydroperoxide substrate) crystals. Data collected for substrate treated oxidised TcGPXI crystals was carried forward for molecular replacement and structural refinement. Values in parentheses refer to the highest resolution shell.

intensities using the program Truncate in the data reduction module of the computer software suite CCP4, where 5 % of the dataset was flagged and assigned as R_{free}.

5.6 Molecular replacement for structure solution of TcGPXI

In order to carry out molecular replacement a starting homology model of wild type TcGPXI was created using the crystal structure of oxidised PtGPX5 (PDB code 2P5R) in the program MODELLER (Chapter 3). The amino acid sequences of wild type TcGPXI and PtGPX5 exhibit a sequence identity of 48 % implying that these proteins are structural homologues. A starting homology model based on the template structure of oxidised PtGPX5 was used for molecular replacement after the loop regions of large sequence diversity and expected structural flexibility in regions of residues Ser46 - Thr51 (containing the peroxidatic cysteine Cys48) and Gly82 - Lys100 (containing postulated resolving cysteine Cys96) were removed. Additionally, residues at the C-terminus and in the loop region of amino acid residues Pro128 - Lys134 (which contains a threonine insertion in wild type TcGPXI (Thr133) otherwise absent in the analogous region of the 2P5R structure), were substituted with alanine or glycine residues or removed. Molecular replacement was carried using the edited starting model and experimental data collected, using the program MOLREP implemented in CCP4. The success of the molecular replacement technique is dependent on the quality of the data, the limits or resolution, the scaling of calculated and observed structure factors and the quality/conformation of the homology model. Rotational and translation functions were calculated using data between 31.47 and 2.81 Å resolution in MOLREP, where the top molecular replacement solution was carried forward for refinement. The top solution from rotation and translation functions implemented in MOLREP was at 87.73° (theta), 16.19° (phi), 52.25° (chi), 0.395 Å (tx), 0.0 Å (ty), 0.218 Å (ty) and had an R_{factor} of 55%.

5.7 Structural refinement of oxidised TcGPXI

The molecular replacement solution was initially refined using cycles of maximum likelihood restrained refinement in REFMAC with intermittent stages of manual fitting, residue building and real-space refinement in the program COOT. Cross-validation was applied using the flagged 5 % of the data (R_{free}). Weighted restraints were applied to various parameters including bond lengths, angles, B-factors, chiral centres and torsion angles. The first cycles of refinement reduced the crystallographic initial R_{factor} and R_{free} values of 0.517 and 0.494 respectively, to 0.368 (R_{factor}) and 0.461 (R_{free}). Further cycles of refinement in REFMAC with application of weighted restraints, manual fitting and model building in COOT facilitated the further reduction of these values.

Residues in loop regions and at the N- and C termini that were initially built as glycine or alanine due to weak electron density were changed to the correct atoms (where possible) according to the primary sequence of TcGPXI as refinement progressed and phases improved. Water molecules were added using cycles of COOT:find waters implemented in REFMAC. Following each stage of refinement weighted $2F_o$-F_c electron density maps were calculated and visualised with the refined model in the program COOT, where atoms were built in, real-space refinement and manual fitting carried out prior to further cycles of refinement. In the final stages of refinement simulated annealing torsion angle dynamics with maximum likelihood function was carried out in CNS (refinement in CNS was kindly carried out by Dr. Snezana Djordjevic), where the temperature of the system was heated to 2500K and slowcooling in appropriate time steps performed. For regions of missing residues, omit maps were calculated in order to carry out appropriate building of residues prior to further cycles of refinement in CNS. Structural refinement using data between resolution range of 2.3 Å and 25.42 Å provided final R_{factor} and R_{free} values of 20.7 % and 27.2 % respectively.

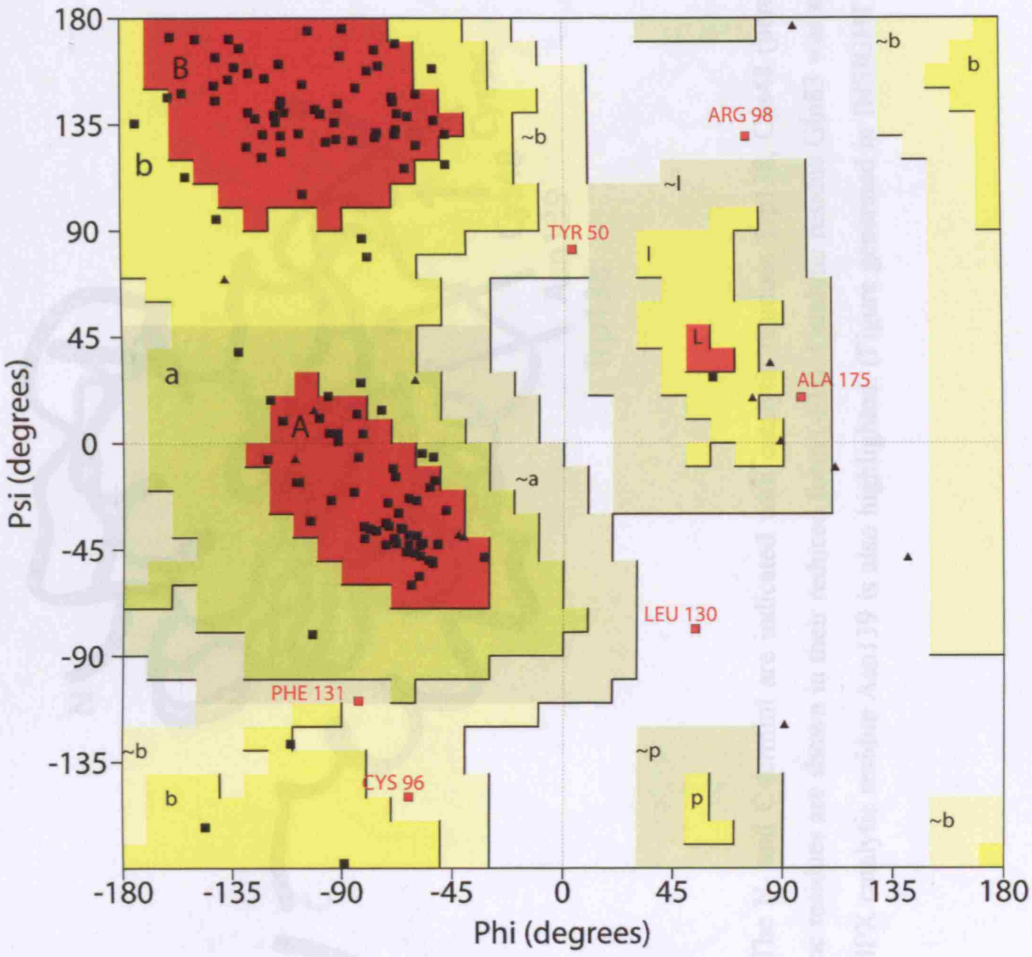
5.8 Structure of oxidised wild type TcGPXI

The structure of wild type TcGPXI was determined from the crystallisation of the protein in its oxidised form, by the treatment of the protein with the substrate

t-butyl hydroperoxide. This eliminated problems in the control of the crystallisation process associated with the reduced form of the protein, which only gave rise to crystals after a period of months (by which time oxidation had occurred) as opposed to a period of days with the substrate-treated protein. The protein was crystallised in the space group C2 where the asymmetric unit contained one molecule of the protein. From the symmetry operators of the designated space group, four molecules were observed in the unit cell. The structure was solved by molecular replacement and refined to a crystallographic R_{factor} and R_{free} of 20.7 % and 27.2 %, respectively. Following iterative stages of refinement and subsequent calculations of electron density maps, all residues were found to be clearly identified in the $2F_o$ - F_c electron density maps except in the region of residues Gln83 - Glu93 (where Gln83 is implemented in catalysis) and for residue Thr133 which were not included in the final model. Additionally, two molecules of glycerol were located resulting from the crystallisation solution and the cryoprotectant. The model also contains 75 water molecules. The overall geometry of the structure was refined to an RMSD of 0.021 Å and 2.031° for bond lengths and bond angles respectively, with an average B-factor of 31.2. The structure was assessed by the program PROCHECK. The Ramachandran plot revealed 82.3 % of residues to be in preferred regions and 16.1 % in allowed regions (Figure 5.9).

5.8.1 Overall fold and secondary structure

The overall fold of the wild type TcGPXI tertiary structure reveals the presence of a globular protein containing a thioredoxin fold that is characteristic of the GPX enzyme family (Figure 5.10). The overall thioredoxin fold of TcGPXI contains mainly four central β strands (of which β5 is antiparallel) flanked by four α helices. In Figure 5.11, the positions of the conserved main catalytic residues, characteristic of the GPX family of enzymes are highlighted in addition to a second redox active cysteine (Cys96) functioning as a resolving cysteine. Despite the absence of catalytic residue Gln83 that was not visible in the electron density, the adjacent residue Gly82 is highlighted. There are additional small secondary structural elements that are analogous to those typically observed in other GPX structures at the N-terminus. A diagrammatic representation of the overall secondary



Plot statistics

Residues in most favoured regions [A,B,L]	102	82.3%
Residues in additional allowed regions [a,b,l,p]	16	12.9%
Residues in generously allowed regions [~a,~b,~l,~p]	4	3.2%
Residues in disallowed regions	2	1.6%
Number of non-glycine and non-proline residues	124	100.0%
Number of end-residues (excl. Gly and Pro)	4	
Number of glycine residues (shown as triangles)	15	
Number of proline residues	9	
Total number of residues	152	

Figure 5.9: Ramachandran plot for the structure of TcGPXI in the oxidised fold

The Ramachandran plot from structural analysis and validation for oxidised TcGPXI structure illustrates the dihedral angles (phi and psi angles) for each amino acid residue. The plot shows that only two residues (Arg98 and Leu130) are found to be in disallowed regions whilst all other residues are found to occur in allowed regions. These two residues are found in loop regions that are only partially built in the TcGPXI structure.

Figure 5.10: Stereo view of TcGPXI in the oxidised fold

Stereo view of oxidised TcGPXI is illustrates in ribbon diagrams. The N- and C-termini are indicated with catalytic residues Trp138, Cys48 (peroxidatic cysteine) and Cys96 (resolving cysteine) shown as sticks. The cysteine residues are shown in their reduced form (-SH). Catalytic residue Gln83 was not built due to poor electron density in this region. The recently postulated GPX catalytic residue Asn139 is also highlighted. (Figure generated in INSIGHT II 98.0 molecular graphics software).

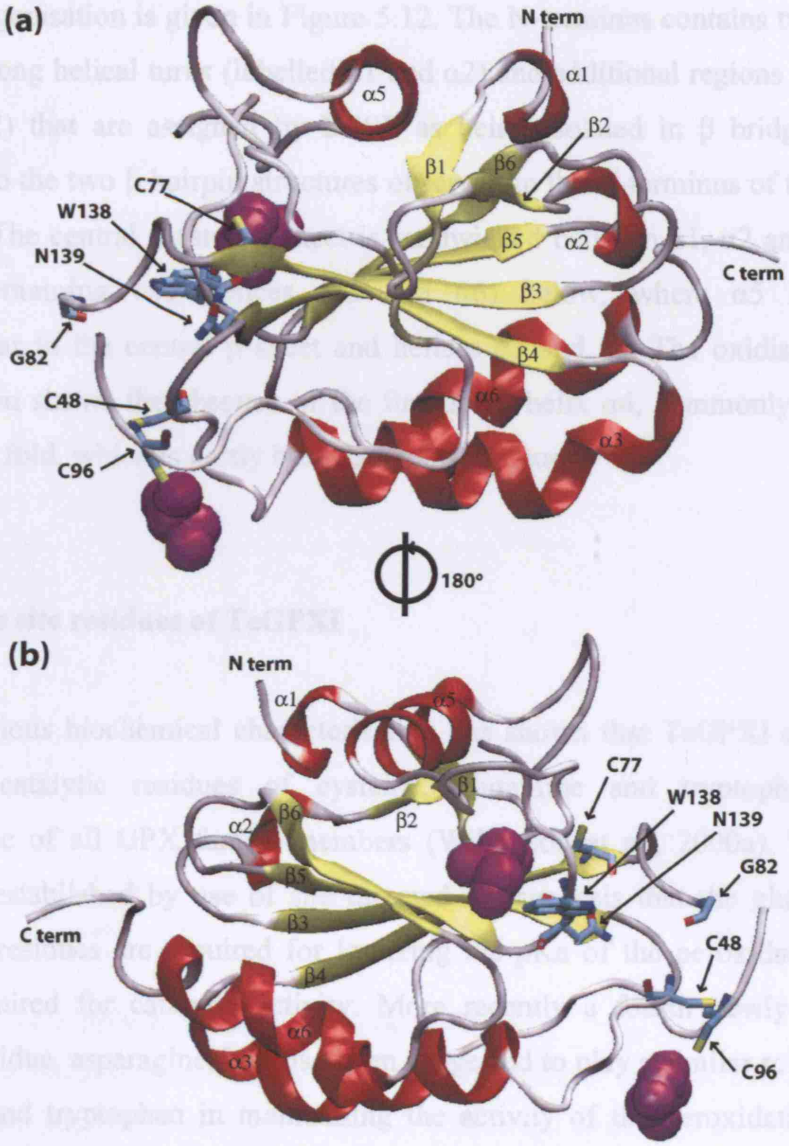


Figure 5.11: Ribbon diagrams of the structure of oxidised TcGPXI

The structure of oxidised TcGPXI is represented as ribbon diagrams where (a) is rotated by 180° to give (b). The structure shows the presence of a thioredoxin fold comprising an α-β-α sandwich, where α-helices (red) flank a central core of β-strands (yellow) with additional shorter secondary structural elements at the N-terminus. Catalytic residues Trp138, Asn139, Cys48 are shown as sticks, in addition to Cys77. Although catalytic residue Gln83 was not identified in the electron density and not built in the model, the position of adjacent residue Gly82 is indicated. The resolving cysteine of TcGPXI (Cys96), occurs in a loop structure that is present in the oxidised fold of GPX-type enzymes exhibiting an atypical 2-Cys peroxiredoxin mechanism, is highlighted. Despite the presence of Cys48 and Cys96 in their dithiol form, evidence strongly suggests that these residues participate in a disulphide bond in the oxidised fold. Two glycerol molecules resulting from the cryoprotectant are shown as surface spheres (purple). (Figure generated in VMD (Humphrey et al., 1996)).

structure organisation is given in Figure 5.12. The N-terminus contains two separate 3 residues-long helical turns (labelled α1 and α2) and additional regions (designated β1 and β 2) that are assigned by DSSP as being isolated in β bridges and are analogous to the two β hairpin structures observed in the N-terminus of the PtGPX5 structures. The central twisted β sheet is sandwiched between α1, α2 and α5 above and the remaining two helices (α3 and α6) below, where α5 lies almost perpendicular to the central β sheet and helices α3 and α6. The oxidised TcGPXI conformation shows the absence of the functional helix α4, commonly found in a thioredoxin fold, which is partly built as a loop structure.

5.8.2 Active site residues of TcGPXI

Previous biochemical characterisation has shown that TcGPXI contains the conserved catalytic residues of cysteine, glutamine and tryptophan that is characteristic of all GPX family members (Wilkinson et al., 2000a). It has been previously established by use of site directed mutagenesis that the glutamine and tryptophan residues are required for lowering the pKa of the peroxidatic cysteine residue required for catalytic activity. More recently a fourth newly postulated catalytic residue, asparagine, has also been suggested to play a similar role to that of glutamine and tryptophan in maintaining the activity of the peroxidatic cysteine. Therefore, in the following section the equivalent catalytic residues of TcGPXI will be considered

The catalytic residue glutamine (Gln83) of TcGPXI is localised in the loop region that is only partially built due to disorder in the electron density around this region and thus Gln83 is absent in this model. Although catalytic residue Gln83 is absent from the model, the surface exposed adjacent residue Gly82 is positioned at a Cα – Cα distance of 9.2 Å from the surface exposed residue Trp138. In accordance with recent work proposing the presence of a catalytic tetrad for the GPX family and additional residues involved, the crystal structure of TcGPXI shows the orientation of the fourth catalytic residue Asn139 which is thought to be necessary for catalysis by involvement in the formation of hydrogen bonds for the deprotonation and orientation of the catalytic cysteine prior to the reduction of the hydroperoxide substrate (Figure 5.11). The close proximity of additional residues Asn43 and Cys77

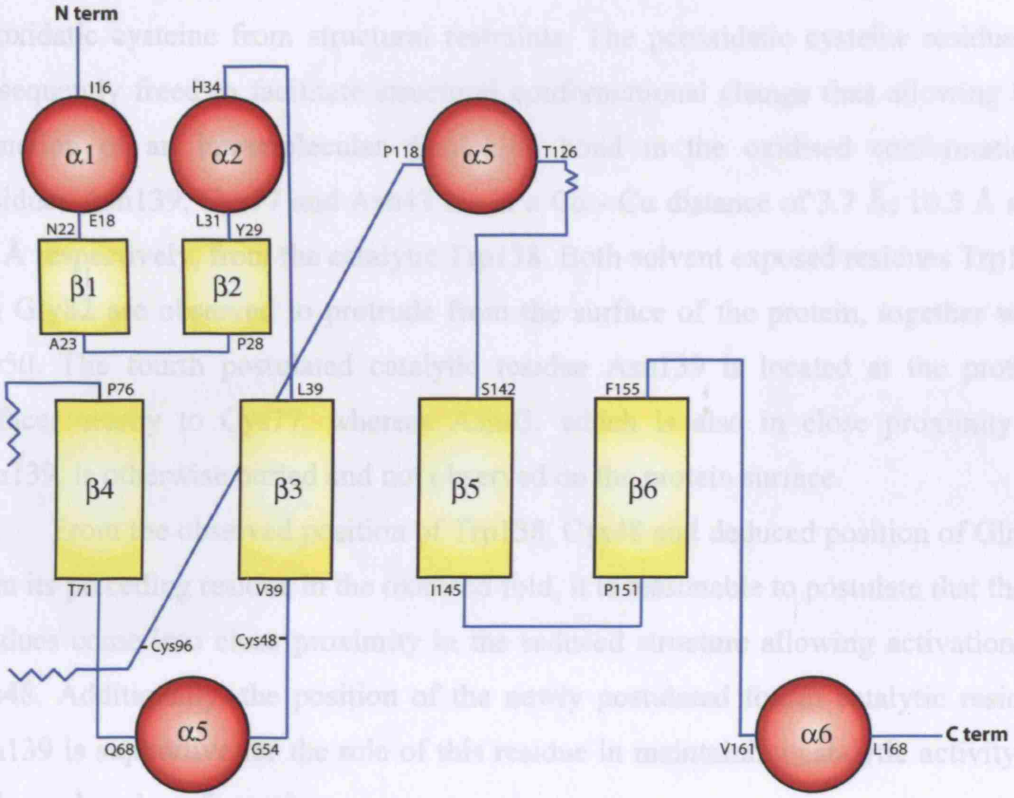


Figure 5.12: Secondary structure of oxidised TcGPXI

Diagrammatic representation of the organisation of secondary structural elements in the oxidised conformation of TcGPXI. The secondary structural elements α-helices (red circles) and β-strands (yellow boxes) are shown connected by loop regions (blue) where regions represented by zig-zag lines indicate loop regions that were not built i.e. Gln83-Glu93 and residue Thr133. Smaller secondary structural elements present at the N-terminus (α1, α2, β1, β2) are also illustrated.

to Asn139 (Cα - Cα distance 5.8 Å and 8.4 Å respectively) is supportive for the postulated role of these two residues in the coordination of catalytic Asn139 following reduction of the hydroperoxide substrate and thus freeing the oxidised peroxidatic cysteine from structural restraints. The peroxidatic cysteine residue is subsequently freed to facilitate structural conformational change thus allowing the formation of an intramolecular disulphide bond in the oxidised conformation. Residues Asn139, Cys77 and Asn43 are at a Cα - Cα distance of 3.7 Å, 10.5 Å and 8.3 Å respectively, from the catalytic Trp138. Both solvent exposed residues Trp138 and Gly82 are observed to protrude from the surface of the protein, together with Tyr50. The fourth postulated catalytic residue Asn139 is located at the protein surface, nearby to Cys77, whereas Asn43, which is also in close proximity to Asn139, is otherwise buried and not observed on the protein surface.

From the observed position of Trp138, Cys48 and deduced position of Gln83 from its preceding residue in the oxidised fold, it is reasonable to postulate that these residues come into close proximity in the reduced structure allowing activation of Cys48. Additionally, the position of the newly postulated fourth catalytic residue Asn139 is supportive for the role of this residue in maintaining catalytic activity in such a reduced conformation.

Upon examination of the electrostatic potential surface map for the structure of oxidised TcGPXI (Figure 5.13a-b), it is clearly observed that one side of the molecule where the redox centre is proposed to occur is predominantly positively charged, whilst the opposite side is predominantly negative in charge. The surface structure exhibits a strongly positively basic charged cavity in which the fourth postulated catalytic residue Asn139 resides, lined by protruding catalytic residue Trp138 and residue Gly82 (adjacent to catalytic Gln83 absent in this structure). The catalytic Trp138 is surrounded by a region of positively charged basic amino acids such as Lys137 and Arg154. The apolar residues Phe80, Ala81, Phe131, Gly132, Ala135, Ile136, Ser156, Pro157 and Gly158 which occur in analogous positions (from sequence alignment) to those residues implicated in the positioning of substrate in reduced human GPX4 (Scheerer et al., 2007), may be involved in the positioning of lipid hydroperoxides at the active site of TcGPXI in its reduced conformation provided a similar orientation of these residues remains in the reduced fold.

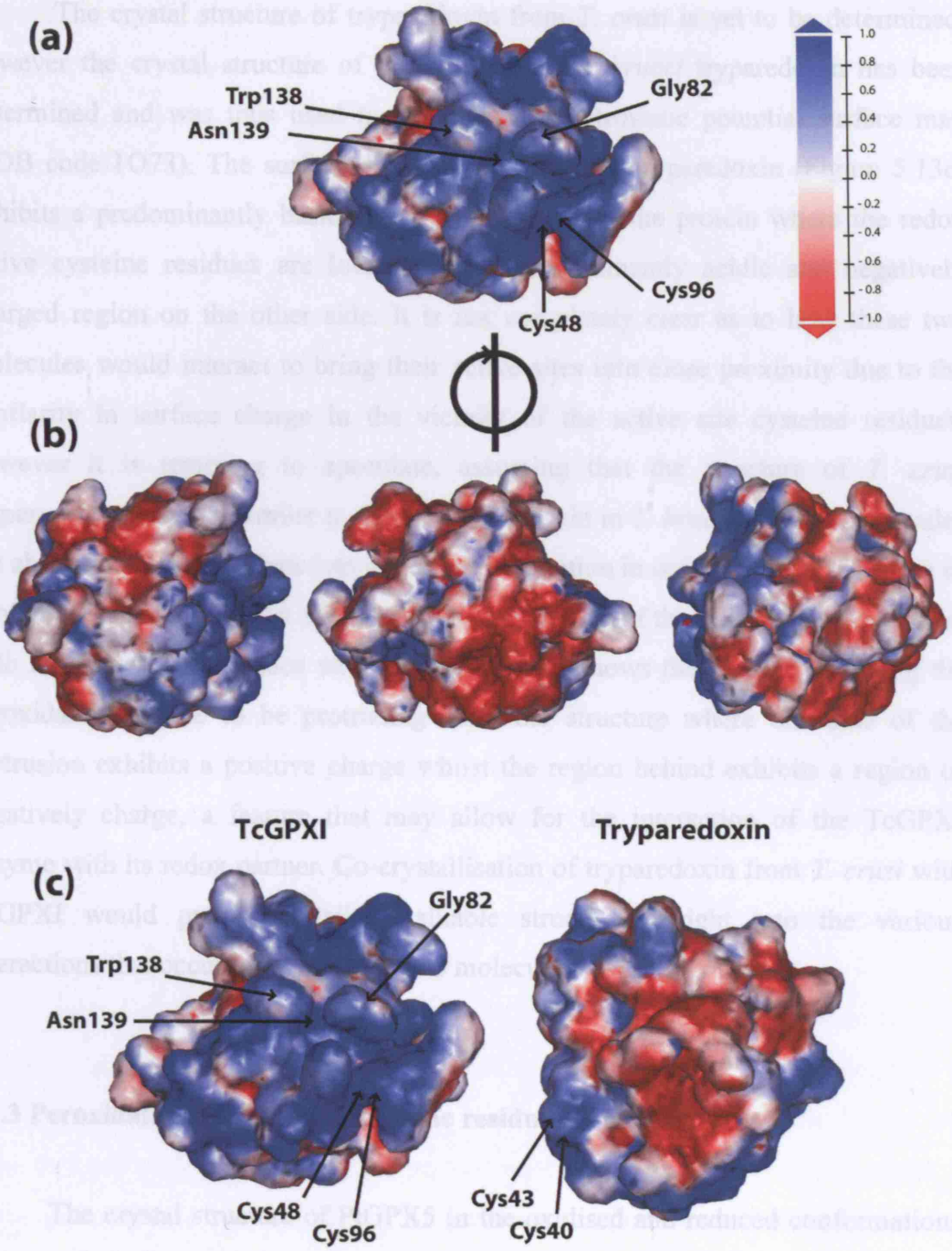


Figure 5.13: Electrostatic-potential surface maps of oxidised TcGPXI

The electrostatic surface potential map of oxidised TcGPXI is shown (a), where blue indicates positively charged regions and red indicates negatively charged regions. Residues postulated to be involved in catalysis: Cys48, Cys96, Trp138, Asn139 are indicated, in addition to residue Gly82 which is adjacent to absent catalytic residue Gln83. Rotation of (a) in 90° steps is shown in (b). The electrostatic surface potential maps of oxidised TcGPXI and *T. brucei* tryparedoxin (PDB code 1O73) are shown (c), where *T. brucei* tryparedoxin redox cysteine residues (Cys40 and Cys43) are indicated. (Figure generated in INSIGHT II 98.0 molecular graphics software).

The crystal structure of tryparedoxin from *T. cruzi* is yet to be determined, however the crystal structure of the homologue *T. brucei* tryparedoxin has been determined and was thus used to generate an electrostatic potential surface map (PDB code 1O73). The surface structure of *T. brucei* tryparedoxin (Figure 5.13c) exhibits a predominantly basic region on one side of the protein where the redox active cysteine residues are located and a predominantly acidic and negatively charged region on the other side. It is not completely clear as to how these two molecules would interact to bring their active sites into close proximity due to the similarity in surface charge in the vicinity of the active site cysteine residues, however it is tempting to speculate, assuming that the structure of *T. cruzi* tryparedoxin is highly similar to that of tryparedoxin in *T. brucei*, that the molecules are able to orient themselves into a feasible orientation in order to avoid repulsion of similar charges faces whilst still facilitating interaction of the redox active sites from both molecules. The surface structure of TcGPXI shows the region containing the peroxidatic cysteine to be protruding from the structure where one side of the protrusion exhibits a positive charge whilst the region behind exhibits a region of negatively charge, a feature that may allow for the interaction of the TcGPXI enzyme with its redox partner. Co-crystallisation of tryparedoxin from *T. cruzi* with TcGPXI would provide detailed valuable structural insight into the various interactions that occur between these two molecules.

5.8.3 Peroxidatic and resolving cysteine residues of TcGPXI

The crystal structure of PtGPX5 in the oxidised and reduced conformations has revealed the presence of a mechanism analogous to the atypical 2-Cys peroxiredoxin enzymes, whereby the oxidised conformation is characterised by the formation of an intramolecular disulphide bond resulting from the unravelling of helix present in the reduced form (containing the resolving cysteine). Koh et al., 2007 reported a decrease in the Cα - Cα distance between the peroxidatic and resolving cysteine residues of 12.1 Å relative to the reduced conformation, associated with a structural changes in the unwinding of the helix containing the resolving cysteine and a conformational change of the loop region containing the

peroxidatic cysteine. In this thesis a similar mechanism is postulated for TcGPXI, which has previously been determined for the homologous *T. brucei* PxIII enzyme.

The crystal structure of wild type TcGPXI is indicative of an oxidised conformation where the catalytic peroxidatic cysteine is found to be in close proximity to the postulated resolving cysteine (Cys96), the latter of which lies close to the two-fold symmetry axis (Figure 5.14a). Both cysteine residues are shown to be in a reduced redox state where despite anticipation of a disulphide bond characteristic of the oxidised conformation, refinement of the structure yielded improved refinement statistics in the absence of a disulphide bond despite the overall conformation being characteristic of an oxidised fold. The phi and psi angles for the peroxidatic cysteine (Cys48) are -108.8 and 132.6 respectively; the phi and psi angles for resolving cysteine (Cys96) are -52.3 and -144.9 respectively. The distance between the SG atoms of Cys48 and Cys96 is 4.16 Å whereas the distance between the SG atom of Cys96 and Cys96' of the neighbouring symmetry related molecule is 3.0 Å. The ideal distance for the formations of a disulphide bond between two thiols is 2.05 ± 0.03 Å (Creighton, 1988), therefore despite the absence of a disulphide bond the close proximity of the resolving and peroxidatic cysteine residues in the structure of oxidised TcGPXI is suggestive that these two residues participate in an intramolecular disulphide bond in the oxidised conformation of TcGPXI.

Although the protein was oxidised by treatment with the substrate and the data is indicative of an oxidised conformation, it is postulated that subsequent to crystal formation, due to the effects of X-ray radiation the disulphide bridge has been reduced within the crystal whilst the two cysteine residues have remained in close proximity to one another. This phenomenon has been observed for other protein crystals for example in the crystals of the DsbG disulphide isomerise protein, where the reduced form of redox cysteine residue is favoured in the equilibrium between disulphide and dithiol forms (Heras et al., 2004). Therefore, it may be concluded from the close proximity of the two cysteine residues that the crystal structure presented in this thesis is representative of an oxidised conformation despite the presence of the peroxidatic and resolving cysteine dithiol groups. Figure 5.14b shows the positions of catalytic residues Cys48, Trp138 and the postulated resolving cysteine, Cys96. Both the side chain of Cys48 and Trp138 is observed to face outwards from the protein core, where specifically the side chain of Cys48 is

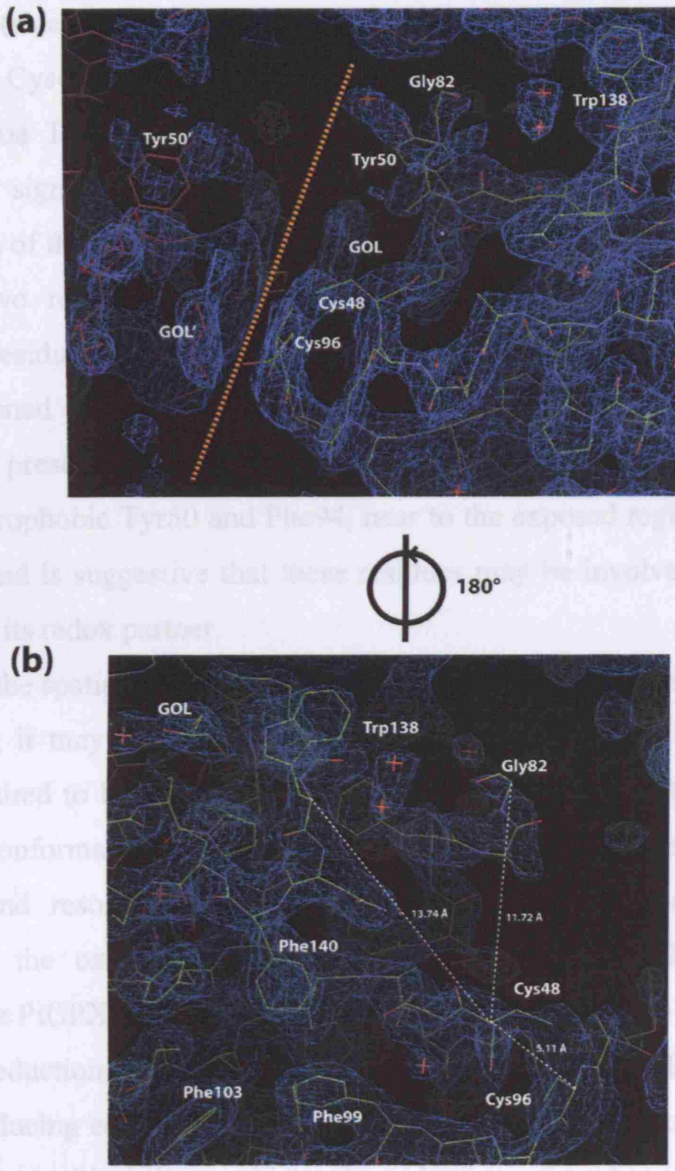


Figure 5.14: Electron density map around the active site residues of oxidised TcGPXI

Electron density for the weighted 2Fo-Fc map in the region of the active site residues is shown, viewed at 0.98 sigma in COOT. Catalytic residue Trp138 and Gly82 (adjacent to catalytic residue Gln83) are shown (a), in addition to a nearby glycerol molecule (GOL). Redox active cysteine residues (Cys48 and Cys96) in their dithiol form are highlighted. The Cys96 residue is close to the two-fold symmetry axis (orange dashed line). Residue Tyr50' and glycerol (GOL') of the symmetry related molecule (pink) are shown. Rotation of (a) by 180 ° is shown in (b). The Cα – Cα distances between: Cys48 and Trp138 (13.74 Å), Cys48 and Gly82 (11.72 Å), Cys48 and Cys96 (5.11 Å) are highlighted showing that the peroxidatic cysteine (Cys48) is closer to the resolving cysteine (Cys96) in the oxidised fold than the region near to Trp138 and Gly82 where catalytic residue Gln83 is speculated to reside contributing to the postulated active site of the reduced form of TcGPXI. The position of the second glycerol molecule near to Trp138 is also shown (b).

facing Cys96 (side chain of Cys96 is perpendicular to that of Cys48). It may be observed that Cys48 is closer to Cys96 (Cα – Cα distance 5.11 Å) rather than catalytic residue Trp138 (Cα – Cα distance 13.74 Å). This further supports the proposal that significant structural differences exist between the reduced and oxidised states of the enzyme.

The two redox active cysteine residues are flanked on either side by hydrophobic residues Tyr50 and Phe94. It should be noted that a glycerol molecule is also positioned nearby to Tyr50 and the surface exposed redox active cysteine residues. The presence of the glycerol molecule and unusual surface exposed side chains of hydrophobic Tyr50 and Phe94, near to the exposed region of the expected disulphide bond is suggestive that these residues may be involved in interaction of TcGPXI with its redox partner.

From the spatial arrangement of the catalytic residues (Cys48 and Trp138) in this structure, it may be speculated that in order for catalysis to occur whereby Cys48 is required to be positioned within hydrogen bonding distance of Trp138 in the reduced conformation of TcGPXI, a structural change occurs thus bringing the peroxidatic and resolving cysteine in close proximity to one another for the formation of the oxidised conformation similar to the mechanism previously established for PtGPX5.

The reduction of the oxidised disulphide form of TcGPXI occurs by the transfer of reducing equivalents from the thioredoxin-like molecule tryparedoxin. A thioredoxin recognition site motif for the reduction of oxidised PtGPX5 has been proposed where the resolving cysteine is located in the sequence motif "EPGxx(D/E)xIxx(F/M)(V/A)CT(R/K)FK" (x denotes any amino acid residue). The primary sequence of TcGPXI exhibits a similar sequence motif "EPGTALEVKEFACTRFK" in which the postulated resolving cysteine (Cys96) is located, which may infer a potential recognition site for tryparedoxin for the subsequent reduction of the disulphide bond in oxidised TcGPXI. However, only the terminal seven residues of this sequence motif were built in the TcGPXI structure, thus improved experimental data with the presence of a fully oxidised form and completion of this loop region might be required to fully analyse the structural features of the loop and to allow inferences on the interaction of oxidised TcGPXI with tryparedoxin.

5.9 Correlation of NMR data to the crystal structure of TcGPXI

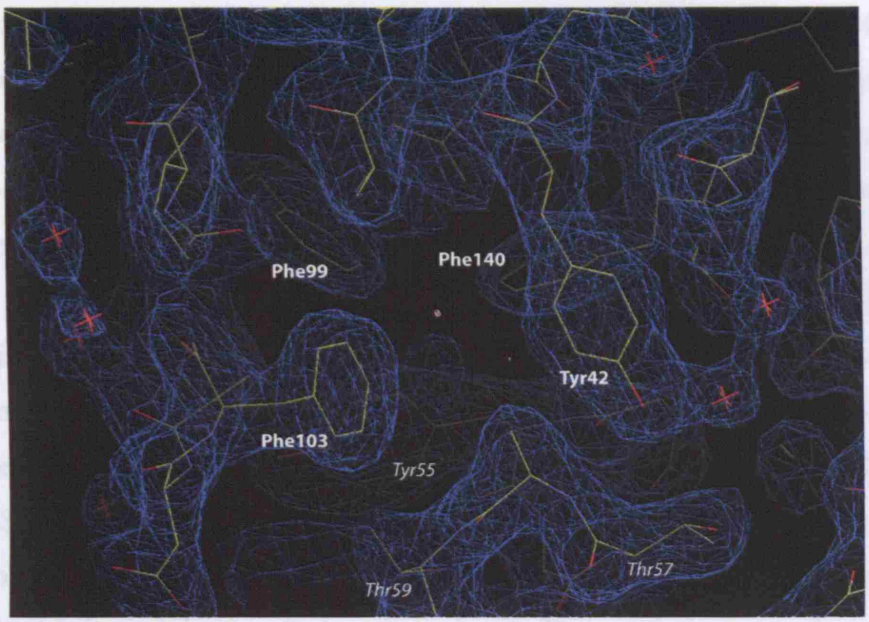
The NMR data for oxidised TcGPXI predicted the presence of a disulphide bond between Cys48 and Cys96, and evidence from the crystal structure data is strongly supportive of this observation. The secondary structure prediction from NMR is also in good agreement with that observed from the X-ray data. Additionally, the structural investigations of oxidised TcGPXI using NMR spectroscopy revealed the observation of unusual amide chemical shifts in the ^{1}H, ^{15}N-HSQC spectrum for residues Leu119, Phe103, Thr71 and Trp138.

Hydrophobic residue Phe103 resides within the loop region containing the resolving cysteine (Cys96) where the aromatic side chain is facing towards the interior of the protein. Interestingly additional aromatic side chains of Phe99, Phe140, Tyr42 and Tyr55 (tyrosine is partially hydrophobic) are also oriented towards each other, with additional residue Val72 and Val44 nearby contributing to the overall hydrophobic character of this region (Figure 5.15a). No apparent unusual features may be observed for Phe103 in the TcGPXI crystal structure that would account for the unusual chemical shift observed through NMR experiments, therefore it may be implied that the unusual chemical shift observed is specifically correlated to the dynamics of the protein in its solution state.

In the crystal structure of TcGPXI, the residue Thr71 is positioned at the beginning of a beta strand (β3), which contributes to the central β-sheet. The unusual chemical shift for Thr71 observed from the NMR investigations may be due to the hydrogen bonding between the amide nitrogen atom of Thr71 and the carbonyl oxygen atom of Pro38 that occurs at the start of neighbouring parallel β2 strand. The distance between the nitrogen atom of Thr71 and oxygen atom of nearby Pro38 is 3.85 Å, which indicates the presence of a weak hydrogen bond. This weak interaction maybe correlated to the fact that Thr71 occurs at the start of the strand rather than towards the middle of the strand where hydrogen bond distances between the backbone NH groups of one strand and neighbouring backbone CO (carbonyl) group of the adjacent strand ranges between approximately 2.5 – 2.8 Å.

Another example of an unusual amide chemical shift observed from NMR investigations was that of residue Leu119, which was observed to be located within hydrogen bonding distance of the nearby side chain of His117 residue (Figure 5.15b). The distance between the backbone nitrogen atom of Leu119 and oxygen

(a)



(b)

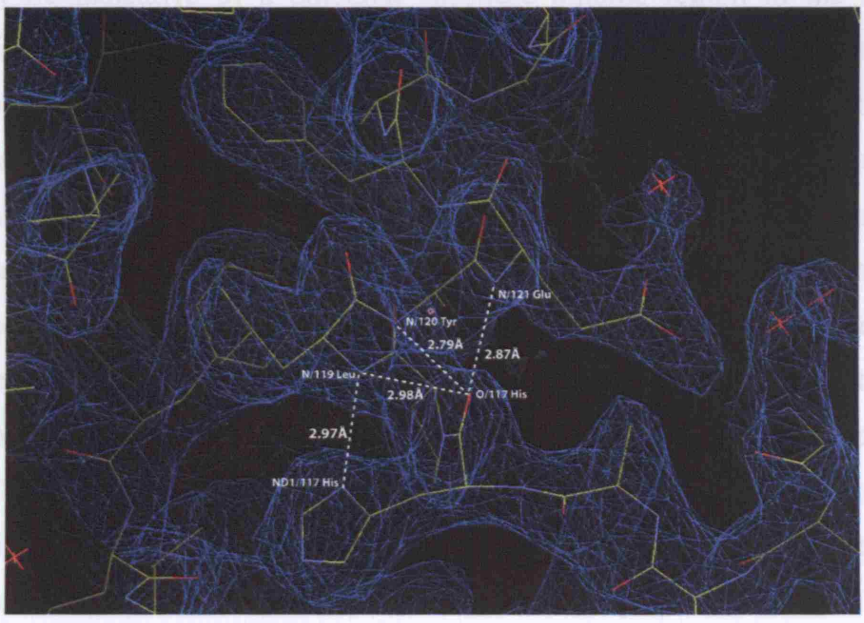


Figure 5.15: Electron density map around the residues Phe103 and Leu119 of oxidised TcGPXI

Electron density for the weighted 2Fo-Fc map in the region of the active site residues is shown, viewed at 0.98 sigma in COOT. The hydrophobic environment of the residue Phe103 is shown (a) where Tyr55, Thr57 and Thr59 form part of α3. Potential hydrogen bonding interactions and atomic distances are illustrated between residue Leu119 and His117 at the start of α5 in oxidised TcGPXI.

atom of His117 is measured to be 2.98 Å. The crystal structure reveals His117, Pro118 and Leu119 to occur as a small three residue turn at the beginning of α5. In protein structures alpha helices usual contain 3.6 residues per turn with hydrogen bonds between the C=O of residue *n* and NH of residue *n* + 4. Due to the absence of an NH group in proline, this residue is not commonly found in α-helices as it would interrupt the hydrogen bonding patterns that regulate the secondary structure and are thus more commonly found at the termini of helices. Often such special sequence conformations act to stabilise the termini of helices in protein structures (Lesk, 2001; Brandon and Tooze, 1999).

The final residue that was found to have unusual amide chemical shifts in the ^{1}H, ^{15}N-HSQC spectrum was catalytic residue Trp138. The crystal structure reveals the hydrophobic tryptophan residue to be surface exposed on a loop region bridging β5 and α5 with its indole ring facing the solvent. The catalytic tryptophan residue contributes to the hydrophobicity from nearby Ile136, Tyr50, Pro157 and Phe80, which may be indicative of a potential substrate binding region in the reduced conformation. Due to the absence of residues Gln83 (catalytic residue) – Glu93, it is not possible to make definitive conclusions on the potential hydrogen bonding patterns that occur between tryptophan in this region which may give rise to unusual chemical shifts observed in the NMR experiments. The completion of the model in this region may provide valuable insight into the structural interactions that may occur in the region of the active site of the oxidised conformation, which give rise to a distinct chemical environment of Trp138 as indicated in NMR experiments. In *E. coli* thioredoxin, a tryptophan residue associated with the active site has been postulated to contribute to the hydrophobic region of the binding site with little conformational change in the residue between the oxidised and reduced state (Holmgren, 1973). The tryptophan residue is known to play an important role in stacking interactions in protein structures through the indole ring as well as being a key residue in the binding sites of many proteins, therefore the fact that the indole ring of Trp138 in TcGPXI is positioned pointing outwards to the solvent in this structure may indicate potential π-π stacking interactions with hydroperoxide substrates (*e.g.* cumene hydroperoxide) of the TcGPXI enzyme (discussed in Chapter 4).

5.10 Glycerol molecule ligands in the structure of oxidised TcGPXI

In order to collect X-ray data from the crystal of oxidised TcGPXI a cryoprotectant of 20 % glycerol was used prior to exposure of the crystal to the X-ray beam. The use of glycerol as a cryoprotectant is common, in addition to its use directly in the crystallisation solution itself whereby glycerol has been found to stabilise conformational flexibility and thus aid crystallisation (Sousa, 1995). From the X-ray data of oxidised TcGPXI regions of electron density in the $2F_o$-F_c map (viewed at 1.18 sigma level) were found correlating to two molecules of glycerol resulting from the cryoprotectant solution (Figure 5.11).

Examination of the crystal structure of oxidised TcGPXI reveals the position of the first glycerol molecule in the surface region behind the catalytic Trp138 (Figure 5.16a). The opposite face of this tryptophan side chain lines the region postulated to comprise the active site in the reduced conformation of TcGPXI together with the peroxidatic cysteine and helix/loop region where catalytic Gln83 is expected to reside. Residues Arg154 (distance of Arg154-NE to O2-glycerol: 2.79 Å), Lys137 (distance of Lys137-N to O3-glycerol: 2.86 Å) and Thr141 (distance of Thr141-OG1 to O3-glycerol: 3.29 Å) are located near to the glycerol molecule with their respective side chains oriented towards the glycerol molecule, suggestive of interactions that contribute to the glycerol binding. Additional polar interactions with the glycerol molecule occur from nearby a water molecule that is at distances 3.37 Å and 2.78 Å from the O1 and O2 atoms respectively, of the glycerol molecule.

The second glycerol molecule is located at the protein surface between the peroxidatic (distance of Cys48-SG to O2-glycerol: 5.48 Å) and resolving cysteine (distance of Cys96-SG to O2-glycerol: 3.73 Å) residues, near Tyr50 (distance of Tyr50-Cα to C3-glycerol: 4.20 Å) whose side chain projects above the resolving and peroxidatic cysteine residues (Figure 5.16b). It may be speculated that the positioning of the glycerol molecule may be indicative of a region where a thiol reductant molecule, such as the known TcGPXI redox partner tryparedoxin, may interact in order to reduce the disulphide bond between the peroxidatic and resolving cysteine residues in a fully oxidised conformation.

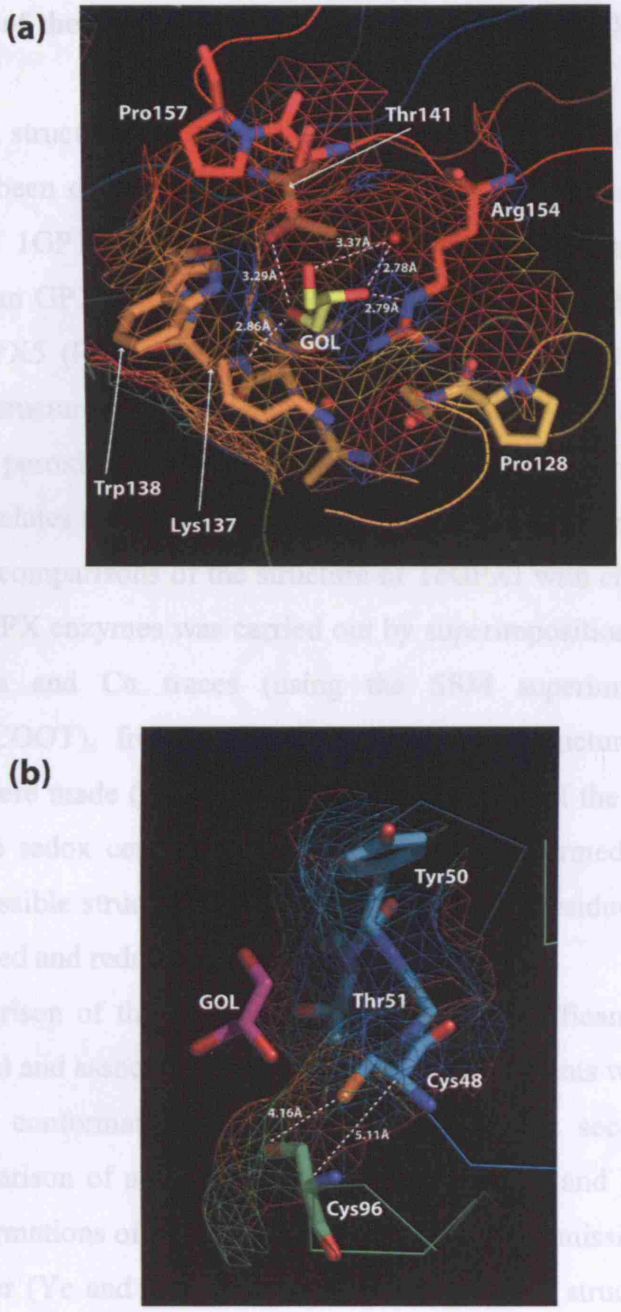


Figure 5.16: Glycerol molecules in the structure of oxidised TcGPXI

Close up view of the two glycerol molecules (GOL) modelled in the structure of oxidised TcGPXI. The first glycerol molecule (a) is found near catalytic Trp138, in a region surrounded by Pro157, Thr141, Arg154, Pro128 and Lys137 (represented as sticks). Polar interactions between the ligand and nearby atoms are illustrated, including to nearby water molecule (red sphere). The second glycerol molecule (b) is found near the peroxidatic (Cys48) and resolving (Cys96) cysteine residues, where Tyr50 and Thr51 are also located (residues shown as sticks). The SG - SG and Cα - Cα atomic distances between the redox active cysteine residues are 4.16 Å and 5.11 Å, respectively. (Figures generated by Pymol (DeLano, W. L., 2002)).

5.11 Comparison of the TcGPXI crystal structure to other GPX homologues

The crystal structures of various GPX family members that are homologues of TcGPXI have been determined: PtGPX5 (PDB code 2P5R and 2P5Q), bovine GPX1 (PDB code 1GP1), human GPX1 (PDB code 2F8A), human GPX2 (PDB code 2HE3), human GPX3 (PDB code 2R37), human GPX4 (PDB code 2GS3 and 2OBI), human GPX5 (PDB code 2I3Y) and human GPX7 (PDB code 2P31). In addition to these structures, most recently the crystal structure of the *T. brucei* GPX-type tryparedoxin peroxidase in the reduced form has been determined (PDB code 2VUP) which correlates to the sequence referred to as TbPxII in this thesis.

Individual comparisons of the structure of TcGPXI with crystal structures of the homologous GPX enzymes was carried out by superimposition of the secondary structure elements and Cα traces (using the SSM superimposition program implemented in COOT), from which observations in structural similarity and variable regions were made (Figure 5.17). An examination of the catalytic residues contributing to the redox centre of each enzyme was performed in order to gain insight into the possible structural changes of the catalytic residues that may occur between the oxidised and reduced conformations of TcGPXI.

The comparison of the GPX structures showed significant conservation of the thioredoxin fold and associated secondary structural elements with only marginal deviations in the conformation at the termini of regular secondary structural elements. A comparison of all structures including TcGPXI and both the oxidised and reduced conformations of PtGPX5 was carried out by submission of coordinates to the POSA server (Ye and Godzik, 2005), which provides structural alignments. The results of the superimposition of all GPX structures compared yielded an RMSD of 1.56 Å over 135 amino acid residues. This highlights the significant conservation of the thioredoxin fold in the GPX family across plant, mammal and trypanosomal members. Figure 5.17a-b shows the superimposition of the GPX crystal structures.

An individual comparison of the TcGPXI structure with each of the other structural homologues showed regions of structural variability were predominantly found in the loop region containing the TcGPXI peroxidatic cysteine (Cys48) between residues Asn43 – Gly53 where certain conformations were found to affect the N- terminal residues of the proceeding helix (α3) causing either tighter or looser

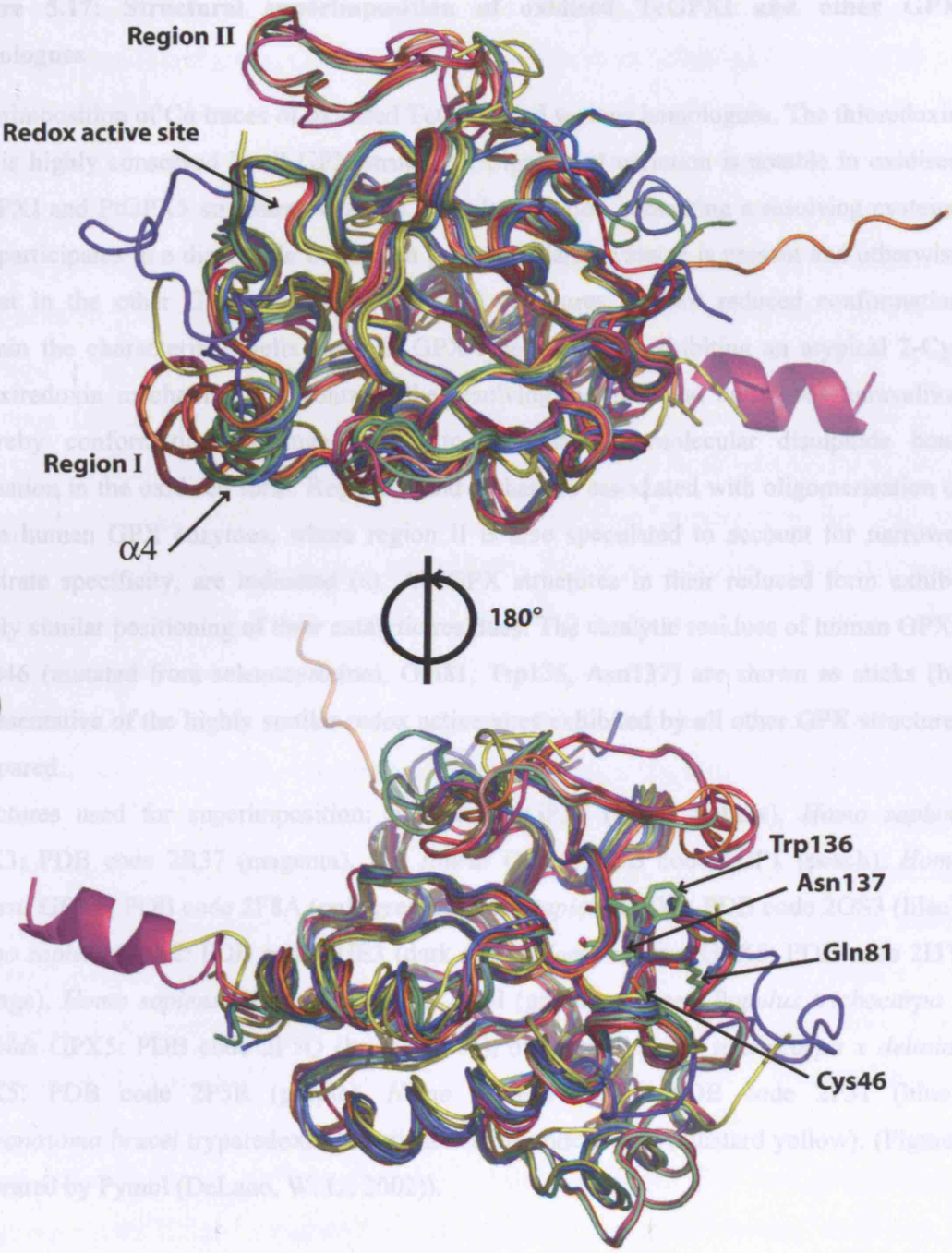


Figure 5.17: Structural superimposition of oxidised TcGPXI and other GPX homologues

(Figure legend overleaf).

Figure 5.17: Structural superimposition of oxidised TcGPXI and other GPX homologues

Superimposition of Cα traces of oxidised TcGPXI and various homologues. The thioredoxin fold is highly conserved in all GPX structures. Significant variation is notable in oxidised TcGPXI and PtGPX5 structures where a large loop region containing a resolving cysteine that participates in a disulphide bond with the peroxidatic cysteine is present and otherwise absent in the other GPX structures. All GPX structures in their reduced conformation contain the characteristic helix (α4). In GPX-type enzymes exhibiting an atypical 2-Cys peroxiredoxin mechanism, α4 contains the resolving cysteine and undergoes unravelling whereby conformational change occurs to facilitate intramolecular disulphide bond formation in the oxidised form. Regions I and II that are associated with oligomerisation of some human GPX enzymes, where region II is also speculated to account for narrower substrate specificity, are indicated (a). All GPX structures in their reduced form exhibit highly similar positioning of their catalytic residues. The catalytic residues of human GPX4 (Cys46 (mutated from selenocysteine), Gln81, Trp136, Asn137) are shown as sticks (b), representative of the highly similar redox active sites exhibited by all other GPX structures compared.

Structures used for superimposition: oxidised TcGPXI (bright yellow), *Homo sapiens* GPX3: PDB code 2R37 (magenta), *Bos taurus* GPX1: PDB code 1GP1 (peach), *Homo sapiens* GPX1: PDB code 2F8A (pale green), *Homo sapiens* GPX4: PDB code 2GS3 (lilac), *Homo sapiens* GPX2: PDB code 2HE3 (dark pink), *Homo sapiens* GPX5: PDB code 2I3Y (orange), *Homo sapiens* GPX4: PDB code 2OBI (green), reduced *Populus trichocarpa x deltoids* GPX5: PDB code 2P5Q (bright green), oxidised *Populus trichocarpa x deltoids* GPX5: PDB code 2P5R (purple), *Homo sapiens* GPX7: PDB code 2P31 (blue), *Trypanosoma brucei* tryparedoxin-peroxidase: PDB code 2VUP (mustard yellow). (Figures generated by Pymol (DeLano, W. L., 2002)).

initial turns only but not carried into the rest of the helix. The second region of variability was found to occur approximately between residues Ala81 – Asp102 of TcGPXI, of which the loop region containing the catalytic residue Gln83 is partially missing in this structure. This sequence region is found to contain the resolving cysteine (Cys96) in TcGPXI, which is thought to participate in a disulphide bond with the peroxidatic cysteine (Cys48) in the oxidised conformation. All structures of GPX homologues that were used for structural comparison except of the structure of oxidised PtGPX (PDB code 2P5R) contain an extra helix (α4), which in those enzymes that are known to exhibit an atypical 2-Cys peroxiredoxin mechanism (e.g. PtGPX5) contains the resolving cysteine that undergoes a local unwinding to bring the peroxidatic and resolving cysteine residues into close proximity in the oxidised form. In addition to the presence of α4, an extra loop region (region I) that is present only in oligomeric GPX forms (Chapter 1) is found at the end of α4 (Figure 5.17a). In the tetrameric human GPX enzymes this region is characterised by the sequence motif "PGGG" and is otherwise absent in PHGPX-like members of the GPX family.

The final region of structural variability amongst the GPX members is in the region between residues Thr126 – Lys134 of the TcGPXI structure, where residue Thr133 was not included in the model as it was not clearly visible in the electron density maps. Interestingly significant deviations are observed in this region where oligomeric GPX structures contain a sequence insertion giving rise to a flexible loop (region II) that conforms above the catalytic tryptophan of each of these structures, the absence of which in PHGPX-like GPX members is suggested to account for their wider substrate specificity than other GPX enzymes and facilitates the access of larger substrates such as complex lipid hydroperoxides. The two loop regions (Regions I and II) are suggested to represent the dimer and trimer interfaces in oligomeric human GPX enzymes (Maiorino et al., 2007).

5.11.1 Catalytic residue positions of GPX homologues

A cross comparison of the active site residues between the structures of the various GPX homologues compared in this study and the structure of TcGPXI was carried out. All GPX structures except for that of oxidised PtGPX5 have been crystallised in their reduced forms. The GPX structures contain the conserved

catalytic residues of cysteine/selenocysteine, glutamine (glutamic acid in PtGPX5) and tryptophan, where in some crystal structures the catalytic cysteine or selenocysteine is mutated to either glycine or cysteine which maybe a requirement for crystallisation to stabilise the redox state of these enzymes. The structures of the reduced conformation all reveal highly similar orientation of the catalytic residues being in close proximity to one another and contributing to the redox active site, where notably the side chains of cysteine/selenocysteine, glutamine/glutamic acid and tryptophan are oriented towards each other within potential hydrogen bonding distance. Figure 5.17b illustrates the Cα backbone trace superimposition of the various GPX homologues where the catalytic residues of human GPX4 (PDB code 2OBI) are representative of the highly conserved orientation of the catalytic residues in the reduced conformations of the TcGPXI homologues. Table 5.2 provides a summary of the Cα - Cα distances between the catalytic residues of the GPX structures used for comparison to the structure of TcGPXI presented in this thesis. In addition to this, the newly postulated fourth conserved catalytic residue asparagine residue is also positioned similarly with its side chain directed towards the catalytic cysteine/selenocysteine within hydrogen bonding distance.

Despite similarity in the backbone region containing the catalytic tryptophan residue, the most notable difference upon comparison of the catalytic tryptophan of each of the GPX structures and that of the TcGPXI structure, is the position of the tryptophan side chain. In all other GPX structures including that of oxidised PtGPX5 the indole ring of catalytic tryptophan is in a position resembling a "lid" of the redox centre, however, in oxidised TcGPXI structure electron density shows that the indole ring of catalytic Trp138 is projecting upwards leaving the redox centre uncovered. This feature may be associated specifically with the oxidised conformation of TcGPXI, as the catalytic tryptophan of PtGPX5 adopts the same 'lid' conformation regardless of oxidation state of catalytic cysteine residues.

The absence of catalytic Gln83 in the TcGPXI crystal structure model does not allow definitive conclusions to be made on the position of this catalytic residue. In all GPX structures compared the position of the peptide backbone containing the preceding residues of catalytic glutamine/glutamic acid residue (region containing the conserved "PCNQ" sequence motif) is nearly identical. A comparison of the position of catalytic Glu79 in oxidised and reduced conformations of PtGPX5 shows

	C_P to Gln	Gln to Trp	Trp to C_P	C_P-C_R
***TcGPXI**	11.72	9.45	13.74	5.11
2R37	5.74	10.20	9.35	-
2I3Y	5.94	10.20	9.67	-
1GP1	6.05	9.69	9.48	-
2F8A	5.73	9.80	9.38	-
2HE3	5.74	10.11	9.86	-
2GS3	6.88	10.16	9.28	-
2OBI	5.80	9.98	9.56	-
2P5Q	5.76	10.61	9.9	16.93
2P5R	13.50	13.79	16.59	4.93
2P31	5.74	10.11	9.89	-
2VUP	5.84	10.16	9.37	16.51

Table 5.2: Cα – Cα distances between residues comprising the catalytic triad in the crystal structures of various TcGPXI homologues

The Cα – Cα distances (Å) between the catalytic residues C_P (denotes position of catalytic cysteine or selenocysteine where selenocysteine is mutated to cysteine or glycine in some structures), glutamine (glutamic acid in *Populus trichocarpa x deltoids*) and tryptophan residues is shown. The Cα – Cα distances (Å) between the peroxidatic (C_P) and resolving (C_R) cysteine residues of those peroxidase enzymes known to function in an atypical 2-Cys type peroxiredoxin mechanism are also provided.

Crystal structures used for analysis: *Homo sapiens* GPX3 (2R37; selenocysteine to glycine mutant), *Homo sapiens* GPX5 (2I3Y), *Bos taurus* GPX1 (1GP1), *Homo sapiens* GPX1 (2F8A; selenocysteine to glycine mutant), *Homo sapiens* GPX2 (2HE3; selenocysteine to cysteine mutant), *Homo sapiens* GPX4 (2GS3; selenocysteine to glycine mutant), *Homo sapiens* GPX4 (2OBI; selenocysteine to cysteine mutant), reduced *Populus trichocarpa x deltoides* GPX5 (2P5Q), oxidised *Populus trichocarpa x deltoides* GPX5 (2P5R), *Homo sapiens* GPX7 (2P31) and reduced *T. brucei* GPX-type peroxidase (2VUP).

*Catalytic residue Gln83 is absent in the TcGPXI structure therefore distance to adjacent Gly82 is given.

that Glu79 moves a distance of 6.9 Å (Cα - Cα distance) upon oxidation. The position of Glu79 in the reduced form is similar to the positions of the catalytic glutamine residues in all other structures of reduced GPX enzymes. Interestingly it may be observed that for the reduced structure of PtGPX5 catalytic Glu79 is oriented within hydrogen bonding distance of the peroxidatic cysteine (coherent with established GPX mechanism) where the side chain projects towards the indole ring of nearby catalytic tryptophan and peroxidatic cysteine sulphur atom. However, in the oxidised conformation of PtGPX5, not only does the position of the backbone slightly change in this region, but the side chain of Glu79 no longer points towards the redox centre and instead faces outwards to the solvent away from the redox centre. Therefore, it may be tentatively speculated, assuming that the catalytic Gln83 of TcGPXI adopts a similar position in the reduced form as for all other reduced GPX structures (including reduced PtGPX5), that Gln83 in TcGPXI undergoes a similar change in position in the transition between its reduced and oxidised states.

5.11.1.1 Redox active cysteine residues of GPX homologues

The native amino acid sequences of the GPX homologues contain either a catalytic cysteine or selenocysteine residue. These enzymes have been crystallised either using their native amino acid sequences or mutation of this catalytic residue to either glycine or cysteine.

The crystal structures of the reduced GPX homologues show their catalytic cysteine/selenocysteine position oriented within hydrogen bonding distance to the other catalytic residues of glutamine/glutamic acid and tryptophan, in a formation that is likely to be a requirement of the redox centre to facilitate substrate reduction. These structures are also characterised by the presence of a helix (α4) that is absent in the crystal structures of oxidised PtGPX5 and TcGPXI. Table 5.3 summarises the Cα - Cα distances between the peroxidatic cysteine of oxidised TcGPXI and analogous residues in the GPX homologues that were superimposed, as well as between resolving cysteine residues where relevant. From the analysis of these distances it could be concluded that the deviations in positioning of the peroxidatic cysteine of oxidised TcGPXI and analogous residue in other reduced enzyme structures is of a similar magnitude, providing a strong indication of the positional

	TcGPXI Cys48 (C_P-C_P)	TcGPXI Cys96 (C_R–C_R)
2R37	5.05	-
2I3Y	5.11	-
1GP1	5.98	-
2F8A	6.10	-
2HE3	5.61	-
2GS3	5.96	-
2OBI	5.83	-
2P5Q	10.36	13.70
2P5R	3.22	6.87
2P31	5.42	
2VUP	5.47	14.13

Table 5.3: Comparison of Cα – Cα distances between peroxidatic cysteine of oxidised TcGPXI and peroxidatic cysteine/selenocysteine of other GPX homologues from structural superimposition

The Cα – Cα distances (Å) between the peroxidatic cysteine (C_P) of TcGPXI (Cys48) and peroxidatic cysteine/selenocysteine position of homologous GPX enzymes from Cα trace superimpositions are provided. The Cα – Cα distances (Å) distances between the resolving cysteine (C_P) of TcGPXI (Cys96) and resolving cysteine residues of reduced and oxidised *Populus trichocarpa x deltoides* GPX5 (2P5Q and 2P5R respectively), and *T. brucei* GPX-type peroxidase (2VUP) are also given.

Crystal structures of TcGPXI homologues compared: *Homo sapiens* GPX3 (2R37; selenocysteine to glycine mutant), *Homo sapiens* GPX5 (2I3Y), *Bos taurus* GPX1 (1GP1), *Homo sapiens* GPX1 (2F8A; selenocysteine to glycine mutant), *Homo sapiens* GPX2 (2HE3; selenocysteine to cysteine mutant), *Homo sapiens* GPX4 (2GS3; selenocysteine to glycine mutant), *Homo sapiens* GPX4 (2OBI; selenocysteine to cysteine mutant), reduced *Populus trichocarpa x deltoides* GPX5 (2P5Q), oxidised *Populus trichocarpa x deltoides* GPX5 (2P5R), *Homo sapiens* GPX7 (2P31) and reduced *T. brucei* GPX-type peroxidase (2VUP).

change that may occur for Cys48 in TcGPXI between the oxidised and reduced form. From structural superimposition it is also notable that the greatest discrepancy between the position of the peroxidatic cysteine residue of TcGPXI and analogous residue occurs with reduced PtGPX5 (10.36 Å) while the closest, most similar position is observed within oxidised PtGPX5 (3.22 Å). This data is consistent with the two structures adopting similar conformations in their oxidised states bringing the two redox cysteine residues close together in order to form a disulphide bridge. The distance between resolving cysteine residues of TcGPXI and reduced PtGPX5 (13.70 Å) is highly similar to that observed for reduced *T. brucei* peroxidase (14.13 Å), suggestive of the extent of positional change the resolving cysteine of TcGPXI would undergo if TcGPXI adopts a similar structure in its reduced form. Notably the distance between the resolving cysteine residues of TcGPXI and oxidised PtGPX5 is significantly smaller (6.87 Å) further supportive of the structural and mechanistic similarity of these two enzymes.

In the reduced state of the PtGPX5 enzyme, the resolving cysteine is localised within the functional helix (α4) where the resolving cysteine is at a Cα- Cα distance of 16.93 Å from the peroxidatic cysteine. Following substrate reduction, the functional helix unravels into a large loop region whereby the peroxidatic and resolving cysteine residues are able to form a disulphide bond in the oxidised state. In the oxidised form of PtGPX5 both peroxidatic cysteine and resolving cysteine residues are displaced by a distance of 9.26 Å and 11.11 Å respectively from their positions in the reduced fold, forming an intramolecular disulphide bond in the oxidised fold where these residues have a Cα - Cα distance of 4.93 Å and their respective sulphur atoms are at a distance of 2.02 Å. From the high degree of sequence and structural similarity between TcGPXI and PtGPX5 it is envisaged that a similar phenomenon is observed in TcGPXI whereby a significant conformational change occurs bringing the peroxidatic and resolving cysteine residues into close proximity. As previously discussed the Cα – Cα distance of the peroxidatic and resolving cysteine residues of TcGPXI is 5.11 Å with their respective sulphur atoms at a distance of 4.16 Å and subsequently are unable to form a disulphide bond perhaps due to the reductive effects of X-ray radiation yielding the dithiol form of the cysteine residues within an otherwise oxidised-like conformation. However, from the similarity in Cα – Cα distances of the peroxidatic and resolving cysteine residues in PtGPX5 and TcGPXI, and taking into account the high degree of

structural and sequence homology between the two enzymes it is highly probable that a disulphide bond between the two cysteine residues has been formed prior to crystallisation.

Additional similarity is observed in the positions of the redox cysteine residues in both oxidised TcGPXI and oxidised PtGPX5 structures- where the redox active cysteine residues are positioned at a significant distance to the catalytic tryptophan and the postulated redox active site of the reduced conformation (Figure 5.18).

5.11.2 Comparison of oxidised TcGPXI to the crystal structure of reduced *T. brucei* tryparedoxin peroxidase

Towards the conclusion of the work presented in this thesis, the crystal structure of a GPX-type *T. brucei* tryparedoxin peroxidase enzyme in the reduced conformation was determined and deposited in the PDB (Figure 5.19a). The *T. brucei* peroxidase enzyme is a close homologue of TcGPXI with a sequence identity of 77%, thus it is highly probable that TcGPXI exhibits a significantly similar reduced conformation to that of its *T. brucei* counterpart. This high sequence identity also allows inference of a similar catalytic mechanism for TcGPXI and further supports the hypothesis of this thesis that TcGPXI operates with an atypical 2-Cys peroxiredoxin type catalytic mechanism.

As shown from kinetic investigation by Schlecker at al., 2007 revealing an atypical 2-Cys peroxiredoxin type mechanism for *T. brucei* peroxidase enzymes, the reduced conformation contains the resolving cysteine in a functional helix (α4) that is speculated to undergo local unwinding and conformational change to facilitate the formation of a disulphide bond in the oxidised fold, similar to the catalytic mechanism of PtGPX5. Superimposition of the Cα backbone traces of *T. brucei* peroxidase and TcGPXI structures reveals an RMSD of 2.06 Å over 150 equivalent positions. There is a high degree of structural similarity and conservation of the thioredoxin fold, with most variation occurring in the analogous loop regions containing the peroxidatic cysteine (between Val44 - Tyr55 of TcGPXI), the resolving cysteine (between Phe80 – Phe103 of TcGPXI) and slight variation in loop conformation between residues Gly129 – Lys134 of TcGPXI. The significant

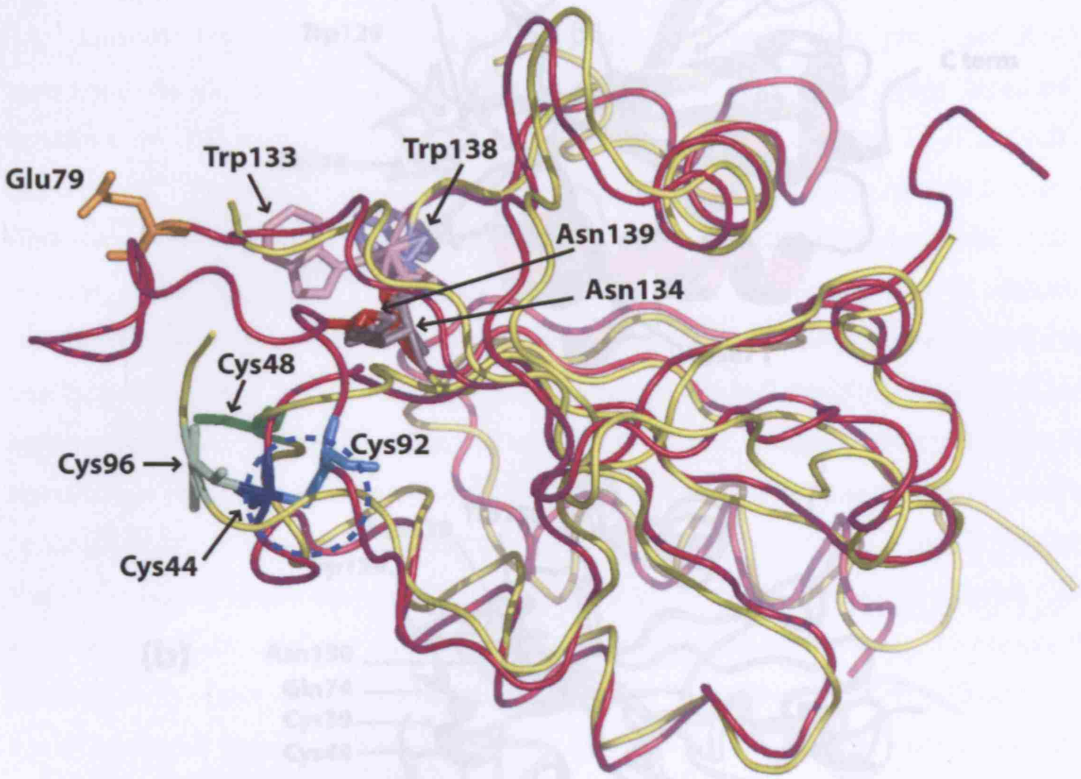


Figure 5.18: Structural superimposition of oxidised TcGPXI and oxidised PtGPX5

The structural superimposition of the Cα traces of oxidised TcGPXI (yellow) and oxidised *Populus trichocarpa x deltoids* GPX5: PDB code 2P5R (PtGPX5; magenta) is depicted in cartoon form. A significant degree of structural conservation is observed, notably by the presence of the thioredoxin fold, with structural deviation most apparent in the loop region containing the intramolecular disulphide bond in oxidised PtGPX5. TcGPXI catalytic residues Trp138 (light blue) and Asn139 (red) are highlighted, in addition to analogous conserved PtGPX5 catalytic residues Trp133 (light pink) and Asn134 (grey). The catalytic residue Glu79 (orange) is shown, of which the analogous catalytic residue (Gln83) is absent in the TcGPXI structure. Peroxidatic cysteine (Cys48; green) and resolving cysteine (Cys96; light green) of TcGPXI are shown in dithiol form although in close proximity. In the oxidised form of PtGPX5 the peroxidatic cysteine (Cys44; blue) and resolving cysteine (Cys92; cyan) participate in an intramolecular disulphide bond (dashed blue circle) that is characteristic of the oxidised conformation. A similar phenomenon is proposed to occur in oxidised TcGPXI. (Figures generated by Pymol (DeLano, W. L., 2002)).

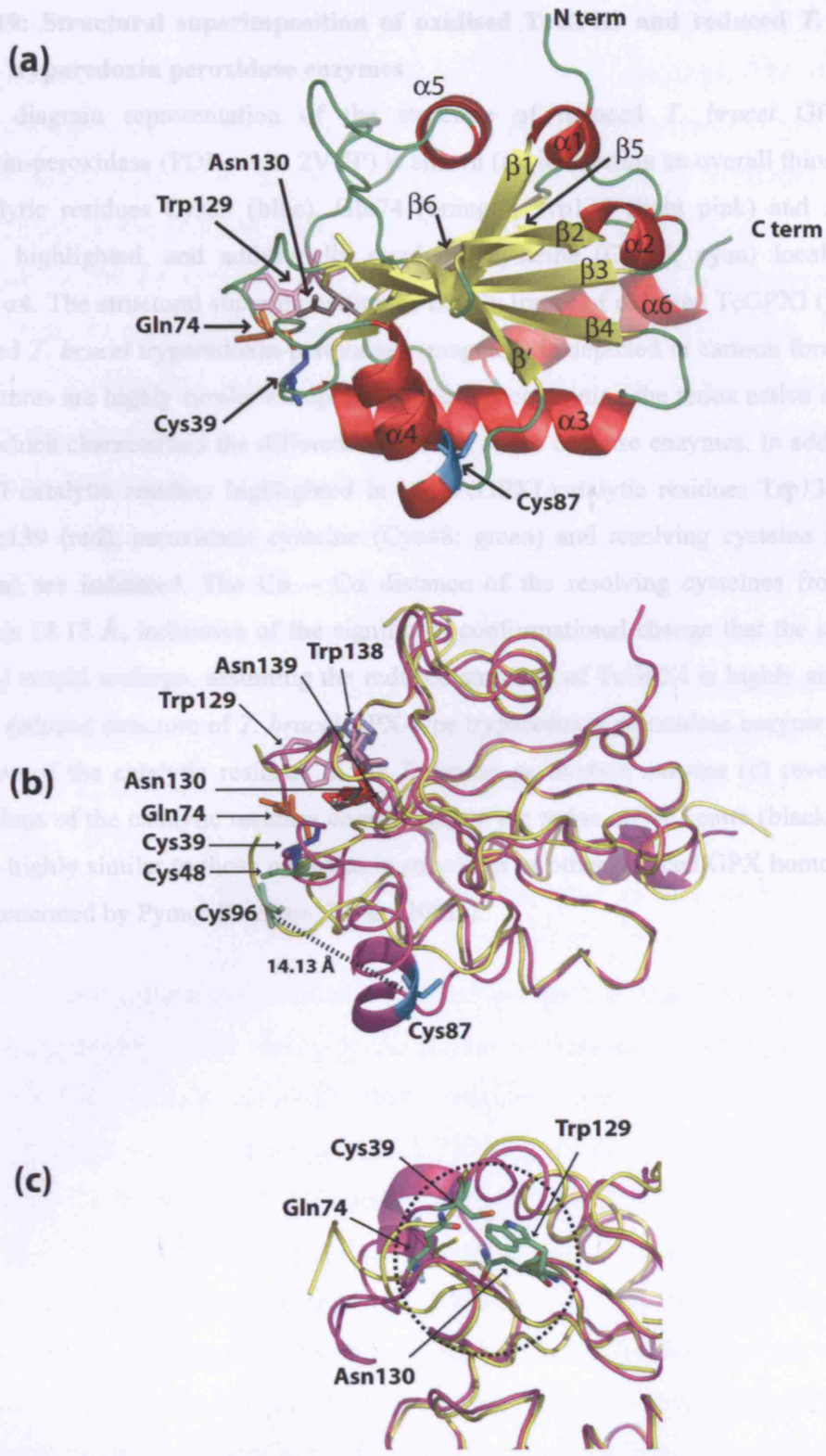


Figure 5.19: Structural superimposition of oxidised TcGPXI and reduced *T. brucei* GPX-type tryparedoxin peroxidase enzymes

(Figure legend overleaf).

Figure 5.19: Structural superimposition of oxidised TcGPXI and reduced *T. brucei* GPX-type tryparedoxin peroxidase enzymes

A ribbon diagram representation of the structure of reduced *T. brucei* GPX-type tryparedoxin-peroxidase (PDB code: 2VUP) is shown (a), illustrating an overall thioredoxin fold. Catalytic residues Cys39 (blue), Gln74 (orange), Trp129 (light pink) and Asn130 (grey) are highlighted, and additionally resolving cysteine (Cys87; cyan) localised in functional α4. The structural superimposition of the Cα traces of oxidised TcGPXI (yellow) and reduced *T. brucei* tryparedoxin-peroxidase (magenta) is depicted in cartoon format (b). Both structures are highly similar except in the regions containing the redox active cysteine residues, which characterises the difference in redox states of these enzymes. In addition to the TbPxII catalytic residues highlighted in (a), TcGPXI catalytic residues Trp138 (light blue), Asn139 (red), peroxidatic cysteine (Cys48; green) and resolving cysteine (Cys96; light green) are indicated. The Cα – Cα distance of the resolving cysteines from both structures is 14.13 Å, indicative of the significant conformational change that the structure of TcGPXI would undergo, assuming the reduced structure of TcGPXI is highly similar to that of the reduced structure of *T. brucei* GPX-type tryparedoxin peroxidase enzyme. Closer examination of the catalytic residues in the *T. brucei* peroxidase enzyme (c) reveals that conformations of the catalytic residues contributing to the redox active centre (black dashed circle) are highly similar to those observed in structures of other reduced GPX homologues. (Figures generated by Pymol (DeLano, W. L., 2002)).

deviations between the structures of the *T. brucei* peroxidase enzyme and TcGPXI, is coherent with the different redox states of these two enzymes. The distance between the resolving cysteine residues of *T. brucei* peroxidase (Cys87) and TcGPXI (Cys96) structures is 14.13 Å, where the former is localised in the functional helix in the reduced state, providing an indication of the conformational change that would be required to facilitate the formation of an intramolecular disulphide bond in these highly homologues enzymes that exhibit an atypical 2-Cys peroxiredoxin mechanism, assuming that both enzymes display similar structures in their difference redox states (Figure 5.19b).

Closer examination of the reduced conformation of the *T. brucei* peroxidase structure reveals an orientation of the catalytic residues that is near identical to the redox site of other GPX structures in their reduced conformation (Figure 5.19c). In the *T. brucei* peroxidase structure the peroxidatic cysteine (Cys39) and resolving cysteine (Cys82) are at a Cα – Cα distance of 16.51 Å. Therefore, accounting for the high sequence similarity between TcGPXI and *T. brucei* peroxidase, it may be speculated that in the reduced conformation the structure of TcGPXI will be highly similar to that of the *T. brucei* peroxidase with a similar redox active site that is characteristic of all the GPX family structures discussed in this section, in their reduced state.

The major difference between *T. brucei* peroxidase and TcGPXI enzymes has been suggested to be the ability of the former to metabolise hydrogen peroxide whereas TcGPXI cannot, although there remains some speculation into this difference. Comparison of the oxidised TcGPXI and reduced TbPxII structures did not reveal significant structural differences around the active site that may account for the difference in substrate specificities between the two enzymes. However, considering that the structures are in different redox states, with significant conformational changes expected to occur in the transition between the two states, it might be more appropriate to examine the structure of TcGPXI in its reduced state to potentially elucidate if and why TcGPXI is unable to metabolise hydrogen peroxide.

5.12 Summary

In order to produce crystals of TcGPXI suitable for X-ray diffraction, purified protein was initially subjected to commercially available crystallisation screens from which suitable conditions were optimised by the production of in-house crystallisation screens. Crystals were found to occur after many months where the crystallisation process was found difficult to control due to the redox nature of the enzyme. In order to control crystallisation the enzyme was oxidised by treatment of TcGPXI with the substrate t-butyl hydroperoxide prior to use with designed in-house crystallisation condition screens. Crystals of substrate treated oxidised TcGPXI suitable for X-ray experiments were produced in conditions of unbuffered 1.8 - 2.0 M ammonium sulphate at 20 °C using the hanging drop vapour-diffusion methods. Untreated native wild type TcGPXI crystals suitable for X-ray diffraction were also produced using hanging drop techniques in conditions of unbuffered 1.8 M ammonium sulphate.

Crystals of substrate treated oxidised wild type TcGPXI and untreated ("native") oxidised TcGPXI were subjected to an X-ray beam, whereby full datasets were collected and processed.

The crystal structure of oxidised wild type TcGPXI was solved in a C2 spacegroup by molecular replacement using an edited homology model based on the template structure of the oxidised conformation of PtGPX5, where loop regions of structural diversity where deleted. The top solution from molecular replacement was carried forward and refined using the CCP4 program REFMAC in conjunction with torsion angle dynamics of the simulated annealing protocol in CNS. Intermittent stages of manual model building and real space refinement in the program COOT were also applied. The structure of oxidised TcGPXI was able to be refined to a crystallographic R_{factor} and R_{free} of 20.7 % and 27 %, respectively.

The crystal structure of oxidised wild type TcGPXI revealed the presence of the highly conserved thioredoxin fold comprising an α-β-α sandwich that is characteristic of the GPX family. Electron density in the $2F_o$-F_c maps corresponding to two glycerol molecules resulting from the cryoprotectant, were identified. The peroxidatic cysteine (Cys48) and resolving cysteine (Cys96) were found to be positioned within close proximity to one another, where despite the existence of these residues in their dithiol form due to the effects of X-ray radiation, the TcGPXI

molecule adopted an oxidised-like fold. This is further supported by the fact that the redox active cysteine residues are at a significant distance from the catalytic tryptophan and speculated position of catalytic residue Gln83. Due to weak electron density in this region we were unable to build a model for residue Gln83. Although, overall the loop containing residues Gln83 and Cys96 which, in a reduced from is expected to form an alpha helix, appeared to be highly disordered the electron-density fro Cys96 was clearly visible.

A comparison of the oxidised TcGPXI structure, homologous human GPX enzymes, the oxidised and reduced conformations of PtGPX5, and the reduced structure of homologous *T. brucei* tryparedoxin peroxidase illustrated a highly conserved thioredoxin fold for all enzymes. The oxidised TcGPXI structure was found to predominantly differ in regions containing its peroxidatic and resolving cysteine residues when compared to reduced conformations of the other GPX enzymes. The main difference in the structures of the reduced GPX enzymes is the presence of a helix (α4), which is otherwise absent in the oxidised conformations of TcGPXI and PtGPX5. In the GPX-type enzymes that exhibit a catalytic mechanism similar to atypical 2-Cys peroxiredoxin enzymes the resolving cysteine is localised in the functional helix (α4) that unwinds as part of a significant conformational change thus facilitating the formation of an intramolecular disulphide bond that is characteristic of the oxidised conformation in these enzymes.

The structure of the TcGPXI homologue *T. brucei* tryparedoxin peroxidase which has been previously established as exhibiting a 2-Cys peroxiredoxin-like catalytic mechanism, does indeed contain a resolving cysteine in a functional helix in its reduced form; thus from the high degree of sequence identity to TcGPXI (sequence identity of 77 %) it is anticipated that a highly similar structure exists for the reduced structure of TcGPXI. In order to fully confirm the structural mechanism of TcGPXI and the accompanied conformational changes that occur, it would be highly valuable to determine the crystal structure in its reduced form as well as to co-crystallise the enzyme with its redox partner tryparedoxin. This would additionally provide insight into the exact interactions that occur between TcGPXI and its redox partner tryparedoxin.

A structural comparison of *T. brucei* tryparedoxin peroxidase and TcGPXI reveals no significant structural differences that may be attributed to the speculated difference in substrate specificities between the two enzymes with respect to

hydrogen peroxide. This further supports the ambiguity that surrounds the ability of TcGPXI to metabolise hydrogen peroxide, where it is possible that catalysis of hydrogen peroxide is carried out by TcGPXI but may be previously undetectable due to effects of overoxidation and hence inactivation of the enzyme.

The determination of the crystal structure of TcGPXI in its oxidised fold provides strong evidence supporting the notion that this enzyme belongs to the class of cysteine-containing GPX-type enzymes from non-mammalian species that exhibit an atypical 2-Cys peroxiredoxin type mechanism. If indeed the structure of the reduced fold of TcGPXI is highly similar to that of the *T. brucei* homologue and their mammalian counterparts, this provides some difficulty in the specific targeting of the parasitic enzyme where specificity problems may only be overcome if significant differences between the parasitic enzyme and its host counterparts are targeted This study has revealed that the most significant difference between the parasitic and host GPX enzymes is the difference in catalytic mechanism as well as the absence of structural features that facilitate oligomerisation in some human GPX isoforms. The formation of an intramolecular disulphide bond in the trypanosomal GPX-type enzymes provides the opportunity for future drug target developments where rational drug design may be focused on the specific targeting of the region containing the redox active cysteine residues in the disulphide bond of the oxidised fold in the parasitic enzymes, thus preventing the regeneration of the reduced fold and hindering parasitic oxidative defence. Additionally the structural determination of TcGPXI may contribute to the potential of inhibiting multiple redox targets in the antioxidant system of trypanosoma in a bid to combat these parasites and their associated diseases in a more efficient method than currently utilised.

Chapter Six

CONCLUSIONS

6.1 Conclusions

The glutathione-dependent enzymes of the antioxidant systems from *T. cruzi* and *T. brucei* have been previously biochemically characterised and have revealed the presence of conserved GPX catalytic residues cysteine, glutamine and tryptophan, in addition to the recently proposed fourth catalytic residue asparagine. While the experiments addressing substrate specificities and subcellular localisations of these enzymes have been carried out, at the beginning of the work described in this thesis there was a lack of structural information that would provide insight into the underlying molecular basis of the catalytic mechanism of these proteins. Unlike the homologous human GPX enzymes that contain a redox active selenocysteine residue, the trypanosomal GPX enzymes contain a redox active cysteine residue (peroxidatic cysteine) in addition to a downstream second cysteine residue (resolving cysteine) in their primary sequences. The crystal structures of the homologue PtGPX5 in its oxidised and reduced redox states have been determined revealing the presence of an atypical 2-Cys peroxiredoxin mechanism, whereby a resolving cysteine is localised in a functional helix (α4) that undergoes unravelling and conformational change to facilitate the formation of an intramolecular disulphide bond with the peroxidatic cysteine in the oxidised state. From the high sequence similarity of TcGPXI to PtGPX5 and from the establishment of a similar mechanism in a homologous *T. brucei* enzyme (TbPxIII), this study provides evidence that a similar catalytic mechanism operates in TcGPXI.

The work presented in this thesis illustrates the use of sub-cloning techniques, NMR spectroscopy and a combination of site-directed mutagenesis and enzyme kinetic assays in unravelling the catalytic mechanism of the antioxidant enzyme *T. cruzi* GPXI, with a final presentation of the first crystal structure of a GPX enzyme from the *T. cruzi* parasite (TcGPXI).

6.2 Cloning, protein purification and crystallisation

In order to produce sufficient and suitable protein for structural studies, various constructs were designed for *T. cruzi* GPXI (TcGPXI), *T. cruzi* GPXII (TcGPXII), *T. brucei* PxI (TbPxI) and *T. brucei* PxII (TbPxII) and sub-cloned. The

constructs were tested for expression in various *E. coli* cell lines from which a suitable host strain was chosen for all subsequent protein expression and purification of the selected constructs. The expressed proteins were able to be successfully purified using a combination of immobilised metal affinity chromatography and size exclusion chromatography for subjection to crystallisation trials (Chapter 3).

6.3 NMR spectroscopy

The NMR investigation of oxidised TcGPXI resulted in a well-resolved ^{1}H, ^{15}N HSQC spectrum from which near complete sequence-specific backbone amide assignments were made (Chapter 4; Figure 4.5). Interestingly the presence of a disulphide bond was observed between the peroxidatic (Cys48) and resolving cysteine (Cys96) residues of TcGPXI, which is considered to have a stabilising effect on the oxidised fold. At this point there was no indication, in literature that disulfide bond formation is involved in the reaction mechanism and our novel finding had to be investigated further. Subsequent NMR spectroscopy carried out on the reduced TcGPXI revealed a significant conformational change where notably the ^{1}H, ^{15}N HSQC spectrum revealed an absence and variable intensities of crosspeaks, which were found to correlate to predicted loop regions associated with the region of the disulphide bond observed in the oxidised conformation (Figure 4.8). To complete the amide assignments of reduced wild type TcGPXI and thus obtain secondary structure predictions, site directed mutagenesis was employed where Cys48 and Cys96 were mutated to glycine residues (C48G and C96G) in a bid to drive the protein into a reduced fold. A third cysteine (Cys77) was also mutated to glycine (C77G) but resulted in an unstable aggregating protein and was thus not carried forward for NMR studies. Both the C48G and C96G mutant proteins ^{1}H, ^{15}N HSQC spectra were found to be more similar to that of oxidised wild type TcGPXI suggesting that the loss of the disulphide bond is insufficient in driving the mutant proteins into a reduced active conformational fold. Ligand binding studies using NMR spectroscopy with the known TcGPXI substrate t-butyl hydroperoxide, were carried out on both oxidised and reduced wild type TcGPXI in addition to C48G and C96G mutant proteins (Section 4.10). From the results it was postulated that the wild type protein is unable to bind its ligand in its oxidised state suggesting that the

enzyme in a reduced active redox state is required for catalysis. Unfortunately the possible phenomenon of chemical exchange in the NMR experiments rendered the experiments probing confirmation of ligand binding in the reduced fold inconclusive. The NMR investigations of the C48G and C96G mutants in the presence of ligand showed no change, however as these mutated proteins exhibit oxidised-like fold it is possible that a fully reduced active conformation is required for ligand binding to occur. In addition, it could be concluded that substrate binding itself is not required for the described conformational change to occur.

6.4 Enzymatic assays

In order to examine the catalytic mechanism of TcGPXI, enzyme assays using tryparedoxin and glutathione were performed, the results of which confirmed that the wild type TcGPXI enzyme was catalytically active, whereas the TcGPXI mutant C48G, C77G and C96G enzymes were rendered inactive (Section 4.21). The lack of activity exhibited by the C77G mutant TcGPXI enzyme suggested a structural role of this residue, which is supported by the 1D ^{1}H NMR spectrum of the protein. Subsequent to this thesis, site directed mutagenesis of Cys77 to an alternative residue has yielded an enzyme exhibiting catalytic activity although the protein still exhibits some aggregation. The lack of activity of C48G and C96G TcGPXI mutant enzymes is highly suggestive that both Cys48 and Cys96 are required for catalysis, favouring the hypothesis for the presence of an atypical 2-Cys peroxiredoxin mechanism in TcGPXI (Figure 4.23).

6.5 Protein crystallisation

All freshly prepared proteins (wild type TcGPXI and cysteine mutated proteins) were subjected to initial crystallisation trials using commercially available screens from which suitable conditions were optimised by the design of in-house crystallisation screens. Crystals of wild type TcGPXI were produced but found to occur only after a period of months. Crystallisation of TcGPXI was difficult to control due to apparent requirement for protein modification (oxidation) prior to

crystallisation, and therefore in order to control this process the protein was oxidised by addition of the substrate t-butyl hydroperoxide prior to use with in-house designed crystallisation screens (Section 3.7). Protein treated in this manner was found to successfully crystallise rapidly in a period of days where crystals of oxidised TcGPXI suitable for X-ray experiments were produced in conditions of unbuffered 1.8 - 2.0 M ammonium sulphate at 20 °C using the hanging drop vapour-diffusion method. Untreated native wild type TcGPXI crystals suitable for X-ray diffraction were also produced using hanging drop techniques in conditions of unbuffered 1.8 M ammonium sulphate however this crystal exhibited poorer diffraction properties.

6.6 Crystal structure of oxidised TcGPXI

The crystal structure of oxidised wild type TcGPXI was solved in a C 2 space group by molecular replacement using a homology model based on the structure of the oxidised PtGPX5, where loop regions were deleted (Chapter 5). The structure of oxidised TcGPXI was refined to a crystallographic R_{factor} and R_{free} of 20.7 % and 27 %, respectively. All residues were built except Thr131 and the large loop region containing the catalytic residue Gln83 due to weak electron density. The crystal structure of oxidised wild type TcGPXI revealed the presence of the highly conserved thioredoxin fold comprising an α-β-α sandwich that is characteristic of the GPX family (Figure 5.11). Additionally two glycerol molecules were identified, resulting from the cryoprotectant. The peroxidatic cysteine (Cys48) and resolving cysteine (Cys96) residues were found to be within close proximity to one another (Figure 5.14), where despite the existence of these residues in their dithiol form suggested to be due to the reducing effects of X-ray radiation, evidence strongly suggests the presence of an intramolecular disulphide bond characteristic of the oxidised fold.

A comparison of the oxidised TcGPXI structure and other homologous GPX enzymes illustrated a highly conserved thioredoxin fold for all enzymes (Section 5.11). Significant variation was found to occur in regions containing the peroxidatic and resolving cysteine residues of the TcGPXI structure when compared to reduced conformations of the other GPX enzymes. Notably, the main difference in the

structures of the reduced GPX enzymes is the presence of an extra helix (α4), which is otherwise absent in the oxidised conformation of TcGPXI (Figure 5.17). In the GPX-type enzymes that exhibit a catalytic mechanism similar to atypical 2-Cys peroxiredoxin enzymes, the resolving cysteine is localised in the functional helix (α4) that unwinds as part of a significant conformational change thus facilitating the formation of an intramolecular disulphide bond that is characteristic of the oxidised conformation in these enzymes. The structural determination of the TcGPXI homologue *T. brucei* tryparedoxin peroxidase, which has been previously established as exhibiting a 2-Cys peroxiredoxin-like catalytic mechanism, has shown a resolving cysteine in a functional helix in its reduced form (Figure 5.19). Therefore, from the high degree of sequence identity to TcGPXI (77 %) it is postulated that a highly similar structure exists for the reduced structure of TcGPXI. In addition to obtaining the crystal structure of TcGPXI in its reduced form, the co-crystallisation of TcGPXI with its redox partner tryparedoxin would provide further valuable insight into the structural interactions that occur during catalysis for regeneration of the ground state enzyme.

6.7 Concluding remarks

In summary, this study provides strong evidence for the presence of an atypical 2-Cys peroxiredoxin catalytic mechanism in TcGPXI, whereby the oxidised conformation is characterised by the presence of an intramolecular disulphide bond. The determination of the crystal structure of TcGPXI in its oxidised fold provides strong evidence supporting the increasing growth of the class of cysteine-containing GPX-type enzymes that exhibit an atypical 2-Cys peroxiredoxin type mechanism. It is envisaged that this novel family of enzymes exhibit similar mechanistic and structural changes during catalysis. Many cysteine-containing GPX enzymes from non-mammalian species have been found to preferentially utilise thioredoxin as a reductant despite structurally resembling the GPX family. Analogously TcGPXI and *T. brucei* peroxidases have been shown to function more efficiently using thioredoxin-like tryparedoxin, thus it may be possible that these enzymes are also able to utilise trypanosomal thioredoxin despite an absence of thioredoxin reductase. Investigation into the structure and catalytic mechanism of TcGPXII, which differs

from TcGPXI in its preferential use of glutathione as a reductant and greater similarity to human GPX4 than its trypanosomal counterparts may provide further knowledge into the evolution of the GPX family in *T. cruzi*.

The unique catalytic mechanism of the trypanosomal GPXI enzyme provides an opportunity for future drug developments. Structural based drug development utilising enzymes commonly target the active site. However, due to the high degree of homology between parasitic and host GPX enzymes, it may be difficult to achieve specificity. Exploitation of the unique trypanosomal catalytic mechanism whereby specifically the intramolecular disulphide bond of the oxidised conformation is targeted may provide a possible route for drug development. Inhibitors have been designed for trypanothione reductase (Hamilton et al., 2005) whereby the compounds designed and synthesised bind in a time-dependent fashion leading to enzyme isomerisation rendering the enzyme catalytically inactive. This concept may be extrapolated to the design of similar inhibitors of trypanosomal GPX enzymes, which would target the disulphide bond-forming cysteine residues prior to causing a similar phenomenon of irreversible inhibition via enzyme isomerisation. Additionally, compounds that would react with the active site thiols might be designed.

Another alternative for drug development utilising trypanosomal GPX enzymes is to target the interaction of the enzymes with their redox partner tryparedoxin. Although protein-protein interactions have been established to be less druggable due to the requirement of covering large surface areas of protein that contain few features which would allow the selective binding of a small molecule (Hubbard, 2006), the distinct surface region of the disulphide bond as shown in the crystal structure of oxidised TcGPXI may provide a target region for selective ligand binding. If the unique structural region containing the intramolecular disulphide bond can be targeted, it may be possible to lock the disulphide bond in place and thus prevent regeneration of the ground state enzyme or prevent interaction with its redox partner tryparedoxin. The ability to target and lock a disulphide bond for drug development strategies is rare; however this may provide a new paradigm in the specific targeting of trypanosomal GPX enzymes. The ability to crystallise *T. cruzi* GPXI is particularly advantageous, as this would enable the use of co-crystallisation or soaking strategies of candidate target compounds.

Drug design and development targeting TcGPXI may be used in conjunction with other drugs inhibiting multiple key redox targets in the parasite's antioxidant defence mechanism. This could therefore lead to a more efficient combat of the infecting *T. cruzi* parasite, and may provide a treatment that is less harmful to the human host than the toxic side effects of current therapies in the treatment of Chagas disease.

Chapter Seven

REFERENCES

Reference List

Adams,P.D., Pannu,N.S., Read,R.J., and Brunger,A.T. (1997). Cross-validated maximum likelihood enhances crystallographic simulated annealing refinement. Proc. Natl. Acad. Sci. U. S. A. *94*, 5018-5023.

Adams,P.D., Pannu,N.S., Read,R.J., and Brunger,A.T. (1999). Extending the limits of molecular replacement through combined simulated annealing and maximum-likelihood refinement. Acta Crystallogr. D. Biol. Crystallogr. *55*, 181-190.

Alaba, K. (2008). Mutation of residue C77 of Trypanosoma cruzi Glutathione Peroxidase I enzyme to observe C77 requirement in enzymatic activity. *Unpublished*

Alphey,M.S., Bond,C.S., Tetaud,E., Fairlamb,A.H., and Hunter,W.N. (2000). The structure of reduced tryparedoxin peroxidase reveals a decamer and insight into reactivity of 2Cys-peroxiredoxins. J. Mol. Biol. *300*, 903-916.

Alphey,M.S., Gabrielsen,M., Micossi,E., Leonard,G.A., McSweeney,S.M., Ravelli,R.B., Tetaud,E., Fairlamb,A.H., Bond,C.S., and Hunter,W.N. (2003). Tryparedoxins from *Crithidia fasciculata* and *Trypanosoma brucei*: photoreduction of the redox disulfide using synchrotron radiation and evidence for a conformational switch implicated in function. J. Biol. Chem. *278*, 25919-25925.

Altschul,S.F., Gish,W., Miller,W., Myers,E.W., and Lipman,D.J. (1990). Basic local alignment search tool. J. Mol. Biol. *215*, 403-410.

Andrade,H.M., Murta,S.M., Chapeaurouge,A., Perales,J., Nirde,P., and Romanha,A.J. (2008). Proteomic Analysis of *Trypanosoma cruzi* Resistance to Benznidazole. J. Proteome. Res. *7* (6), 2357–2367.

Bateman,A., Birney,E., Durbin,R., Eddy,S.R., Howe,K.L., and Sonnhammer,E.L. (2000). The Pfam protein families database. Nucleic Acids Res. *28*, 263-266.

Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.F., Jr., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T., and Tasumi,M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. J. Mol. Biol. *112*, 535-542.

Berriman,M., Ghedin,E., Hertz-Fowler,C., Blandin,G., Renauld,H., Bartholomeu,D.C., Lennard,N.J., Caler,E., Hamlin,N.E., Haas,B., Bohme,U., Hannick,L., Aslett,M.A., Shallom,J., Marcello,L., Hou,L., Wickstead,B., Alsmark,U.C., Arrowsmith,C., Atkin,R.J., Barron,A.J., Bringaud,F., Brooks,K., Carrington,M., Cherevach,I., Chillingworth,T.J., Churcher,C., Clark,L.N., Corton,C.H., Cronin,A., Davies,R.M., Doggett,J., Djikeng,A., Feldblyum,T., Field,M.C., Fraser,A., Goodhead,I., Hance,Z., Harper,D., Harris,B.R., Hauser,H., Hostetler,J., Ivens,A., Jagels,K., Johnson,D., Johnson,J., Jones,K., Kerhornou,A.X., Koo,H., Larke,N., Landfear,S., Larkin,C., Leech,V., Line,A., Lord,A., Macleod,A., Mooney,P.J., Moule,S., Martin,D.M., Morgan,G.W., Mungall,K., Norbertczak,H., Ormond,D., Pai,G., Peacock,C.S., Peterson,J., Quail,M.A., Rabbinowitsch,E., Rajandream,M.A., Reitter,C., Salzberg,S.L., Sanders,M., Schobel,S., Sharp,S., Simmonds,M., Simpson,A.J., Tallon,L., Turner,C.M., Tait,A., Tivey,A.R., Van,A.S., Walker,D., Wanless,D., Wang,S., White,B., White,O., Whitehead,S., Woodward,J., Wortman,J., Adams,M.D., Embley,T.M., Gull,K., Ullu,E., Barry,J.D., Fairlamb,A.H., Opperdoes,F., Barrell,B.G., Donelson,J.E., Hall,N., Fraser,C.M., Melville,S.E., and El-Sayed,N.M. (2005). The genome of the African trypanosome *Trypanosoma brucei*. Science *309*, 416-422.

Bond,C.S., Zhang,Y., Berriman,M., Cunningham,M.L., Fairlamb,A.H., and Hunter,W.N. (1999). Crystal structure of *Trypanosoma cruzi* trypanothione reductase in complex with trypanothione, and the structure-based discovery of new natural product inhibitors. Structure. *7*, 81-89.

Brandon,C. and Tooze,J. (1999). Introduction to Protein Structure. Garland Publishing Inc.

Brigelius-Flohe,R., Aumann,K.D., Blocker,H., Gross,G., Kiess,M., Kloppel,K.D., Maiorino,M., Roveri,A., Schuckelt,R., and Usani,F. (1994). Phospholipid-

hydroperoxide glutathione peroxidase. Genomic DNA, cDNA, and deduced amino acid sequence. J. Biol. Chem. *269*, 7342-7348.

Brunger,A.T. and Adams,P.D. (2002). Molecular dynamics applied to X-ray structure refinement. Acc. Chem. Res. *35*, 404-412.

Brunger,A.T., Adams,P.D., and Rice,L.M. (1998). Recent developments for the efficient crystallographic refinement of macromolecular structures. Curr. Opin. Struct. Biol. *8*, 606-611.

Burchmore,R.J., Ogbunude,P.O., Enanga,B., and Barrett,M.P. (2002). Chemotherapy of human African trypanosomiasis. Curr. Pharm. Des *8*, 256-267.

Castro,H. and Tomas,A.M. (2008). Peroxidases of Trypanosomatids. Antioxid. Redox. Signal.

Chayen,N.E. (2004). Turning protein crystallisation from an art into a science. Curr. Opin. Struct. Biol. *14*, 577-583.

Chayen,N.E. and Saridakis,E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. Nat. Methods *5*, 147-153.

Clarkson,J. and Campbell,I.D. (2003). Studies of protein-ligand interactions by NMR. Biochem. Soc. Trans. 31, 1006-1009.

Comini,M.A., Krauth-Siegel,R.L., and Flohe,L. (2007). Depletion of the thioredoxin homologue tryparedoxin impairs antioxidative defence in African trypanosomes. Biochem. J. *402*, 43-49.

Creighton,T.E. (1988). Disulphide bonds and protein stability. Bioessays *8*, 57-63.

Declercq,J.P., Evrard,C., Clippe,A., Stricht,D.V., Bernard,A., and Knoops,B. (2001). Crystal structure of human peroxiredoxin 5, a novel type of mammalian peroxiredoxin at 1.5 + resolution. Journal of Molecular Biology *311*, 751-759.

DeLano, W. L. The PyMOL Molecular Graphics System. DeLano Scientific, Palo Alto, CA, USA . 2002.

Delaglio,F., Grzesiek,S., Vuister,G.W., Zhu,G., Pfeifer,J., and Bax,A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR *6*, 277–293

Derome,A.E. (1987). Modern NMR Techniques for Chemistry Research. Pergamon Press).

Division of Parasitic Diseases Centre for Disease Control and Prevention: http://www.cdc.gov/ncidod/dpd/index.htm

Docampo,R. and Moreno,S.N. (2003). Current chemotherapy of human African trypanosomiasis. Parasitol. Res. *90 Supp 1*, S10-S13.

El-Sayed,N.M., Myler,P.J., Bartholomeu,D.C., Nilsson,D., Aggarwal,G., Tran,A.N., Ghedin,E., Worthey,E.A., Delcher,A.L., Blandin,G., Westenberger,S.J., Caler,E., Cerqueira,G.C., Branche,C., Haas,B., Anupama,A., Arner,E., Aslund,L., Attipoe,P., Bontempi,E., Bringaud,F., Burton,P., Cadag,E., Campbell,D.A., Carrington,M., Crabtree,J., Darban,H., da Silveira,J.F., de,J.P., Edwards,K., Englund,P.T., Fazelina,G., Feldblyum,T., Ferella,M., Frasch,A.C., Gull,K., Horn,D., Hou,L., Huang,Y., Kindlund,E., Klingbeil,M., Kluge,S., Koo,H., Lacerda,D., Levin,M.J., Lorenzi,H., Louie,T., Machado,C.R., McCulloch,R., McKenna,A., Mizuno,Y., Mottram,J.C., Nelson,S., Ochaya,S., Osoegawa,K., Pai,G., Parsons,M., Pentony,M., Pettersson,U., Pop,M., Ramirez,J.L., Rinta,J., Robertson,L., Salzberg,S.L., Sanchez,D.O., Seyler,A., Sharma,R., Shetty,J., Simpson,A.J., Sisk,E., Tammi,M.T., Tarleton,R., Teixeira,S., Van,A.S., Vogt,C., Ward,P.N., Wickstead,B., Wortman,J., White,O., Fraser,C.M., Stuart,K.D., and Andersson,B. (2005a). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science *309*, 409-415.

El-Sayed,N.M., Myler,P.J., Blandin,G., Berriman,M., Crabtree,J., Aggarwal,G., Caler,E., Renauld,H., Worthey,E.A., Hertz-Fowler,C., Ghedin,E., Peacock,C.,

Bartholomeu,D.C., Haas,B.J., Tran,A.N., Wortman,J.R., Alsmark,U.C., Angiuoli,S., Anupama,A., Badger,J., Bringaud,F., Cadag,E., Carlton,J.M., Cerqueira,G.C., Creasy,T., Delcher,A.L., Djikeng,A., Embley,T.M., Hauser,C., Ivens,A.C., Kummerfeld,S.K., Pereira-Leal,J.B., Nilsson,D., Peterson,J., Salzberg,S.L., Shallom,J., Silva,J.C., Sundaram,J., Westenberger,S., White,O., Melville,S.E., Donelson,J.E., Andersson,B., Stuart,K.D., and Hall,N. (2005b). Comparative genomics of trypanosomatid parasitic protozoa. Science *309*, 404-409.

Emsley,P. and Cowtan,K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D. Biol. Crystallogr. *60*, 2126-2132.

Epp,O., Ladenstein,R., and Wendel,A. (1983). The refined structure of the selenoenzyme glutathione peroxidase at 0.2-nm resolution. Eur. J. Biochem. *133*, 51-69.

Eswar,N., Webb,B., Marti-Renom,M.A., Madhusudhan,M.S., Eramian,D., Shen,M.Y., Pieper,U., and Sali,A. (2007). Comparative protein structure modeling using MODELLER. Curr. Protoc. Protein Sci *Chapter 2*, Unit.

Fairlamb,A.H. and Cerami,A. (1992). Metabolism and functions of trypanothione in the Kinetoplastida. Annu. Rev. Microbiol. *46*, 695-729.

Falquet,L., Pagni,M., Bucher,P., Hulo,N., Sigrist,C.J., Hofmann,K., and Bairoch,A. (2002). The PROSITE database, its status in 2002. Nucleic Acids Res. *30*, 235-238.

Ferguson,A.D., Labunskyy,V.M., Fomenko,D.E., Arac,D., Chelliah,Y., Amezcua,C.A., Rizo,J., Gladyshev,V.N., and Deisenhofer,J. (2006). NMR Structures of the Selenoproteins Sep15 and SelM Reveal Redox Activity of a New Thioredoxin-like Family. J. Biol. Chem. 281, 3536-3543.

Fernandes,A.P. and Holmgren,A. (2004). Glutaredoxins: glutathione-dependent redox enzymes with functions far beyond a simple thioredoxin backup system. Antioxid. Redox. Signal. *6*, 63-74.

Flohe,L., Budde,H., and Hofmann,B. (2003). Peroxiredoxins in antioxidant defense and redox regulation. Biofactors *19*, 3-10.

Flohe,L. (1998). The Achilles' heel of trypanosomatids: trypanothione-mediated hydroperoxide metabolism. Biofactors *8*, 87-91.

Gasteiger,E., Gattiker,A., Hoogland,C., Ivanyi,I., Appel,R.D., and Bairoch,A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. Nucleic Acids Res. *31*, 3784-3788.

Gommel,D.U., Nogoceke,E., Morr,M., Kiess,M., Kalisz,H.M., and Flohe,L. (1997). Catalytic characteristics of tryparedoxin. Eur. J. Biochem. *248*, 913-918.

Grzesiek,S. and Bax,A. (1993). Amino acid type determination in the sequential assignment procedure of uniformly 13C/15N-enriched proteins. J. Biomol. NMR *3*, 185–204.

Hamilton, C.J., Saravanamuthu, A.,Poupta, C., Fairlamb, A.H., Eggleston, I.M. (2005). Time-dependent inhibitors of trypanothione reductase : Analogues of the spermidine alkaloid lunarine and related natural products. Bioorganic & medicinal chemistry *14*,2266-2278.

Hannaert,V., Saavedra,E., Duffieux,F., Szikora,J.P., Rigden,D.J., Michels,P.A.M., and Opperdoes,F.R. (2003). Plant-like traits associated with metabolism of Trypanosoma parasites. Proceedings of the National Academy of Sciences of the United States of America *100*, 1067-1071.

Harris,R. and Driscoll,P.C. (2007). The Ubiquitin NMR Resource. In Modern NMR spectroscopy, D.Rovnyak and R.A.Stockland, eds. American Chemical Society.

Heras,B.a., Edeling,M.A., Schirra,H.J., Raina,S., and Martin,J.L. (2004). Crystal structures of the DsbG disulfide isomerase reveal an unstable disulfide. Proceedings of the National Academy of Sciences of the United States of America *101*, 8876-8881.

Herbette,S., Roeckel-Drevet,P., and Drevet,J.R. (2007). Seleno-independent glutathione peroxidases. More than simple antioxidant scavengers. FEBS J. *274*, 2163-2180.

Hillebrand,H., Schmidt,A., and Krauth-Siegel,R.L. (2003). A second class of peroxidases linked to the trypanothione metabolism. J. Biol. Chem. *278*, 6809-6815.

Holmgren,A. (1985). Thioredoxin. Annu. Rev. Biochem. *54*, 237-271.

Holmgren,A. (1973). Effects of Oxidation of Tryptophan Residues in Thioredoxin from *Escherichia coli* by N-Bromosuccinimide. J. Biol. Chem. *248*, 4106-4111.

Hubbard, R. E. (2006). Structure-based Drug Discovery. RSC Publishing.

Humphrey, W., Dalke, A., Schulten, K. (1996). VMD - Visual Molecular Dynamics. J. Molec. Graphics. *14*, 33-38.

Hussain, S., Khan, A., Brocklehurst, K. (2002). A review of developments in the study of cysteine proteinase mechansim: opportunities for investigation of electrostatic effects and dynamic aspects of molecular recognition in enzyme chemistry. Recent Res. Devel. Biochem. *3*, 653-677.

Ivens,A.C., Peacock,C.S., Worthey,E.A., Murphy,L., Aggarwal,G., Berriman,M., Sisk,E., Rajandream,M.A., Adlem,E., Aert,R., Anupama,A., Apostolou,Z., Attipoe,P., Bason,N., Bauser,C., Beck,A., Beverley,S.M., Bianchettin,G., Borzym,K., Bothe,G., Bruschi,C.V., Collins,M., Cadag,E., Ciarloni,L., Clayton,C., Coulson,R.M., Cronin,A., Cruz,A.K., Davies,R.M., De,G.J., Dobson,D.E., Duesterhoeft,A., Fazelina,G., Fosker,N., Frasch,A.C., Fraser,A., Fuchs,M., Gabel,C., Goble,A., Goffeau,A., Harris,D., Hertz-Fowler,C., Hilbert,H., Horn,D., Huang,Y., Klages,S., Knights,A., Kube,M., Larke,N., Litvin,L., Lord,A., Louie,T., Marra,M., Masuy,D., Matthews,K., Michaeli,S., Mottram,J.C., Muller-Auer,S., Munden,H., Nelson,S., Norbertczak,H., Oliver,K., O'neil,S., Pentony,M., Pohl,T.M., Price,C., Purnelle,B., Quail,M.A., Rabbinowitsch,E., Reinhardt,R., Rieger,M., Rinta,J., Robben,J., Robertson,L., Ruiz,J.C., Rutter,S., Saunders,D., Schafer,M., Schein,J.,

Schwartz,D.C., Seeger,K., Seyler,A., Sharp,S., Shin,H., Sivam,D., Squares,R., Squares,S., Tosato,V., Vogt,C., Volckaert,G., Wambutt,R., Warren,T., Wedler,H., Woodward,J., Zhou,S., Zimmermann,W., Smith,D.F., Blackwell,J.M., Stuart,K.D., Barrell,B., and Myler,P.J. (2005). The genome of the kinetoplastid parasite, *Leishmania major*. Science *309*, 436-442.

INSIGHT II 98.0 molecular graphics software (Accelrys, San Diego, CA, USA).

Jung,B.G., Lee,K.O., Lee,S.S., Chi,Y.H., Jang,H.H., Kang,S.S., Lee,K., Lim,D., Yoon,S.C., Yun,D.J., Inoue,Y., Cho,M.J., and Lee,S.Y. (2002). A Chinese cabbage cDNA with high sequence identity to phospholipid hydroperoxide glutathione peroxidases encodes a novel isoform of thioredoxin-dependent peroxidase. J. Biol. Chem. *277*, 12572-12578.

Kabsch,W. and Sander,C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers *22*, 2577-2637.

Kamata,H. and Hirata,H. (1999). Redox regulation of cellular signalling. Cell Signal. *11*, 1-14.

Kanelis,V., Forman-Kay,J.D., and Kay,L.E. (2001). Multidimensional NMR methods for protein structure determination. IUBMB. Life 52, 291-302.

Karplus,P.A. and Schulz,G.E. (1989). Substrate binding and catalysis by glutathione reductase as derived from refined enzyme: Substrate crystal structures at 2+ resolution. Journal of Molecular Biology *210*, 163-180.

Katz L.A. and Bhattacharya D. Genomics and Evolution of Microbial Eukaryotes. 2006. Oxford University Press.

Kelly, J. M. Molecular Mechanisms of Pathogenesis in Chagas Disease. 2003. Eurekah.com and Kluwer, Academic/Plenum.

Kissinger,J.C. (2006). A tale of three genomes: the kinetoplastids have arrived. Trends in Parasitology *22*, 240-243.

Krissinel, E., Henrick, K. (2004). Secondary structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta. Crystallogr. *D60*, 2256-2268.

Koh,C.S., Didierjean,C., Navrot,N., Panjikar,S., Mulliert,G., Rouhier,N., Jacquot,J.P., Aubry,A., Shawkataly,O., and Corbier,C. (2007). Crystal structures of a poplar thioredoxin peroxidase that exhibits the structure of glutathione peroxidases: insights into redox-driven conformational changes. J. Mol. Biol. *370*, 512-529.

Kraulis,P.J. (1989). ANSIG: A Program for the Assignment of Protein 1H 2D NMR spectra by Interactive Graphics. Journal of Magnetic Resonance 24, 627-633.

Kundrot,C.E. (2004). Which strategy for a protein crystallization project? Cell Mol. Life Sci. *61*, 525-536.

Larkin,M.A., Blackshields,G., Brown,N.P., Chenna,R., McGettigan,P.A., McWilliam,H., Valentin,F., Wallace,I.M., Wilm,A., Lopez,R., Thompson,J.D., Gibson,T.J., and Higgins,D.G. (2007). Clustal W and Clustal X version 2.0. Bioinformatics. *23*, 2947-2948.

Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. *PROCHECK: A program to check the stereochemical quality of protein structures*. J.Appl.Cryst. *26*, 283-291. 1993.

László Polgár (1989). Mechanisms of Protease Action. CRC Press Inc.

Lesk,A.M. (2001). Introduction to Protein Architecture. Oxford University Press.

Logan,F.J., Taylor,M.C., Wilkinson,S.R., Kaur,H., Kelly,J.M. (2007). The terminal step in vitamin C biosynthesis in *Trypanosoma cruzi* is mediated by a FMN-dependent galactonolactone oxidase. Biochem. J. *407*, 419-426.

Lopez,J.A., Carvalho,T.U., de,S.W., Flohe,L., Guerrero,S.A., Montemartini,M., Kalisz,H.M., Nogoceke,E., Singh,M., Alves,M.J., and Colli,W. (2000). Evidence for a trypanothione-dependent peroxidase system in *Trypanosoma cruzi*. Free Radic. Biol. Med. *28*, 767-772.

Machado,C.R., ugusto-Pinto,L., McCulloch,R., and Teixeira,S.M.R. (2006). DNA metabolism and genetic diversity in Trypanosomes. Mutation Research/Reviews in Mutation Research *612*, 40-57.

MacLeod,A., Tait,A., and Turner,C.M. (2001). The Population Genetics of Trypanosoma brucei and the Origin of Human Infectivity. Philosophical Transactions: Biological Sciences *356*, 1035-1044.

Maiorino,M., Aumann,K.D., Brigelius-Flohe,R., Doria,D., van den,H.J., McCarthy,J., Roveri,A., Ursini,F., and Flohe,L. (1998). Probing the presumed catalytic triad of a selenium-containing peroxidase by mutational analysis. Z. Ernahrungswiss. *37 Suppl 1*, 118-121.

Maiorino,M., Ursini,F., Bosello,V., Toppo,S., Tosatto,S.C.E., Mauri,P., Becker,K., Roveri,A., Bulato,C., Benazzi,L., De Palma,A., and Flohq,L. (2007). The Thioredoxin Specificity of Drosophila GPx: A Paradigm for a Peroxiredoxin-like Mechanism of many Glutathione Peroxidases. Journal of Molecular Biology *365*, 1033-1046.

Marti-Renom,M.A., Stuart,A.C., Fiser,A., Sanchez,R., Melo,F., and Sali,A. (2000). Comparative protein structure modeling of genes and genomes. Annu. Rev. Biophys. Biomol. Struct. *29*, 291-325.

Matthews BW. Solvent content of protein crystals. J.Mol.Biol. 33, 491-497. 1968.

Mielke,S.P. and Krishnan,V.V. (2004). An evaluation of chemical shift index-based secondary structure determination in proteins: influence of random coil chemical shifts. J. Biomol. NMR 30, 143-153.

Miller,J.B. (2000). Why NMR is attracting drug designers. Todays Chemist at Work 9, 44-46.

Muller,S., Liebau,E., Walter,R.D., and Krauth-Siegel,R.L. (2003). Thiol-based redox metabolism of protozoan parasites. Trends Parasitol. *19*, 320-328.

Murshudov,G.N., Vagin,A.A., and Dodson,E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D. Biol. Crystallogr. *53*, 240-255.

Murta,S.M.F., Gazzinelli,R.T., Brener,Z., and Romanha,A.J. (1998). Molecular characterization of susceptible and naturally resistant strains of *Trypanosoma cruzi* to benznidazole and nifurtimox. Molecular and Biochemical Parasitology *93*, 203-214.

Mustacich,D. and Powis,G. (2000). Thioredoxin reductase. Biochem. J. *346 Pt 1*, 1-8.

Nok,A.J. (2003). Arsenicals (melarsoprol), pentamidine and suramin in the treatment of human African trypanosomiasis. Parasitol. Res. *90*, 71-79.

Nordberg,J. and Arnqr,E.S.J. (2001). Reactive oxygen species, antioxidants, and the mammalian thioredoxin system. Free Radical Biology and Medicine *31*, 1287-1312.

Orengo,C.A., Bray,J.E., Buchan,D.W., Harrison,A., Lee,D., Pearl,F.M., Sillitoe,I., Todd,A.E., and Thornton,J.M. (2002). The CATH protein family database: a resource for structural and functional annotation of genomes. Proteomics. *2*, 11-21.

Ottaviano,F.G., Handy,D.E., and Loscalzo,J. (2008). Redox regulation in the extracellular environment. Circ. J. *72*, 1-16.

Oza,S.L., Tetaud,E., Ariyanayagam,M.R., Warnon,S.S., and Fairlamb,A.H. (2002). A single enzyme catalyses formation of Trypanothione from glutathione and spermidine in *Trypanosoma cruzi*. J. Biol. Chem. *277*, 35853-35861.

Pai,E.F. and Schulz,G.E. (1983). The catalytic mechanism of glutathione reductase as derived from x-ray diffraction analyses of reaction intermediates. J. Biol. Chem. *258*, 1752-1757.

Pearl,F., Todd,A., Sillitoe,I., Dibley,M., Redfern,O., Lewis,T., Bennett,C., Marsden,R., Grant,A., Lee,D., Akpor,A., Maibaum,M., Harrison,A., Dallman,T., Reeves,G., Diboun,I., Addou,S., Lise,S., Johnston,C., Sillero,A., Thornton,J., and Orengo,C. (2005). The CATH Domain Structure Database and related resources Gene3D and DHS provide comprehensive domain family information for genome analysis. Nucleic Acids Res. *33*, D247-D251.

Pflugrath,J.W. (1999). The finer things in X-ray diffraction data collection. Acta Crystallogr. D. Biol. Crystallogr. *55*, 1718-1725.

Piattoni,C.V., Blancato,V.S., Miglietta,H., Iglesias,A.A., and Guerrero,S.A. (2006). On the occurrence of thioredoxin in *Trypanosoma cruzi*. Acta Trop. *97*, 151-160.

Pineyro,M.D., Pizarro,J.C., Lema,F., Pritsch,O., Cayota,A., Bentley,G.A., and Robello,C. (2005). Crystal structure of the tryparedoxin peroxidase from the human parasite *Trypanosoma cruzi*. J. Struct. Biol. *150*, 11-22.

Ploubidou,A., Robinson,D.R., Docherty,R.C., Ogbadoyi,E.O., and Gull,K. (1999). Evidence for novel cell cycle checkpoints in trypanosomes: kinetoplast segregation and cytokinesis in the absence of mitosis. J. Cell Sci. *112 (Pt 24)*, 4641-4650.

Raoult,D. (1993). Antimicrobial Agents and Intracellular Pathogens. CRC Press Inc, p. 328.

Rattle,H. (1995). An NMR Primer for Life Scientists. Partnership Press.

Read, R. 1997-2007: Protein Crystallography Course. University of Cambridge

Rhodes, G. Crystallography Made Crystal Clear. third edition. 2006. Academic Press, Elsevier.

Rice,L.M. and Brunger,A.T. (1994). Torsion angle dynamics: reduced variable conformational sampling enhances crystallographic structure refinement. Proteins *19*, 277-290.

Rivera,M.T., De Souza,A.P., Moreno,A.H., Xavier,S.S., Gomes,J.A., Rocha,M.O., Correa-Oliveira,R., Neve,J., Vanderpas,J., and raujo-Jorge,T.C. (2002). Progressive Chagas' cardiomyopathy is associated with low selenium levels. Am. J. Trop. Med. Hyg. *66*, 706-712.

Rossmann, M. G. and Blow, D. M. (1962). The Detection of Sub-Units Within the Crystallographic Asymmetric Unit. Acta Crystallogr. *15*, 24-31.

Rossmann, M. G. (1990). The Molecular Replacement Method. Acta Crystallogr. *46*, 73-82.

Ruegg,U.T. and Rudinger,J. (1977). Reductive cleavage of cysteine disulfides with tributylphosphine. Methods Enzymol. 47, 111-116.

Sanders,J.K.M. and Hunter,B.K. (1987). Modern NMR spectroscopy, a guide for chemists. Oxford University Press.

Scheerer,P., Borchert,A., Krauss,N., Wessner,H., Gerth,C., Hohne,W., and Kuhn,H. (2007). Structural Basis for Catalytic Activity and Enzyme Polymerization of Phospholipid Hydroperoxide Glutathione Peroxidase-4 (GPx4). Biochemistry *46*, 9041-9049.

Schlecker,T., Comini,M.A., Melchers,J., Ruppert,T., and Krauth-Siegel,R.L. (2007). Catalytic mechanism of the glutathione peroxidase-type tryparedoxin peroxidase of *Trypanosoma brucei*. Biochem. J. *405*, 445-454.

Schlecker,T., Schmidt,A., Dirdjaja,N., Voncken,F., Clayton,C., and Krauth-Siegel,R.L. (2005). Substrate specificity, localization, and essential role of the glutathione peroxidase-type tryparedoxin peroxidases in *Trypanosoma brucei*. J. Biol. Chem. *280*, 14385-14394.

Schmidt,H. and Krauth-Siegel,R.L. (2003). Functional and Physicochemical Characterization of the Thioredoxin System in *Trypanosoma brucei*. J. Biol. Chem. *278*, 46329-46336.

Seo,M.S., Kang,S.W., Kim,K., Baines,I.C., Lee,T.H., and Rhee,S.G. (2000). Identification of a New Type of Mammalian Peroxiredoxin That Forms an Intramolecular Disulfide as a Reaction Intermediate. J. Biol. Chem. *275*, 20346-20354.

Sharma,D. and Rajarathnam,K. (2000). 13C NMR chemical shifts can predict disulfide bond formation. Journal of Biomolecular NMR 18, 165-171.

Simpson,A.G.B., Lukes,J., and Roger,A.J. (2002). The evolutionary history of kinetoplastids and their kinetoplasts. Molecular Biology and Evolution *19*, 2071-2083.

Sousa,R. (1995). Use of glycerol, polyols and other protein structure stabilizing agents in protein crystallization. Acta Crystallogr. D Biol. Crystallogr. *51*, 271-277.

Spera,S. and Bax,A. (1991). Empirical correlation between protein backbone conformation and Ca and Cß 13C nuclear magnetic resonance chemical shifts. Journal of the American Chemical Society *113*, 5490-5492.

Sternberg,M.J. and Zvelebil,M.J. (1990). Prediction of protein structure from sequence. Eur. J. Cancer *26*, 1163-1166.

Stuart,K., Brun,R., Croft,S., Fairlamb,A., Gurtler,R.E., McKerrow,J., Reed,S., and Tarleton,R. (2008). Kinetoplastids: related protozoan pathogens, different diseases. Journal of Clinical Investigation *118*, 1301-1310.

Teixeira,A.R., Nitz,N., Guimaro,M.C., Gomes,C., and Santos-Buch,C.A. (2006). Chagas disease. Postgrad. Med. J. *82*, 788-798.

Tetaud,E., Giroud,C., Prescott,A.R., Parkin,D.W., Baltz,D., Biteau,N., Baltz,T., and Fairlamb,A.H. (2001). Molecular characterisation of mitochondrial and cytosolic trypanothione-dependent tryparedoxin peroxidases in *Trypanosoma brucei*. Molecular and Biochemical Parasitology *116*, 171-183.

Toppo,S., Vanin,S., Bosello,V., and Tosatto,S.C. (2008). Evolutionary and Structural Insights Into the Multifaceted Glutathione Peroxidase (Gpx) Superfamily. Antioxid. Redox. Signal. *10 (9)*: 1501-1514.

Tosatto,S.C., Bosello,V., Fogolari,F., Mauri,P., Roveri,A., Toppo,S., Flohe,L., Ursini,F., and Maiorino,M. (2008). The Catalytic Site of Glutathione Peroxidases. Antioxid. Redox. Signal. *10(9)*: 1515-1526.

Turrens,J.F. (2004). Oxidative stress and antioxidant defenses: a target for the treatment of diseases caused by parasitic protozoa. Mol. Aspects Med. *25*, 211-220.

Wang,C., Chen,J., Yin,S., and Chuang,W. (2006). Predicting the redox state and secondary structure of cysteine residues in proteins using NMR chemical shifts. Proteins: Structure, Function and Bioinformatics *63*, 219-226.

Wang,C. and Palmer,A.G. (2003). Solution NMR methods for quantitative identification of chemical exchange in ^{15}N-labelled proteins. Magnetic Resonance in Chemistry *41*, 866-876.

Watson,W.H., Yang,X., Choi,Y.E., Jones,D.P., and Kehrer,J.P. (2004). Thioredoxin and Its Role in Toxicology. Toxicol. Sci. *78*, 3-14.

Wen,J.J., Yachelini,P.C., Sembaj,A., Manzur,R.E., and Garg,N.J. (2006). Increased oxidative stress is correlated with mitochondrial dysfunction in chagasic patients. Free Radic. Biol. Med. *41*, 270-276.

Wilkinson,S.R. and Kelly,J.M. (2003). The role of glutathione peroxidases in trypanosomatids. Biol. Chem. *384*, 517-525.

Wilkinson,S.R., Meyer,D.J., and Kelly,J.M. (2000a). Biochemical characterization of a trypanosome enzyme with glutathione-dependent peroxidase activity. Biochem. J. *352 Pt 3*, 755-761.

Wilkinson,S.R., Meyer,D.J., Taylor,M.C., Bromley,E.V., Miles,M.A., and Kelly,J.M. (2002a). The *Trypanosoma cruzi* enzyme TcGPXI is a glycosomal peroxidase and can be linked to trypanothione reduction by glutathione or tryparedoxin. J. Biol. Chem. *277*, 17062-17071.

Wilkinson,S.R., Prathalingam,S.R., Taylor,M.C., Horn,D., and Kelly,J.M. (2005). Vitamin C biosynthesis in trypanosomes: a role for the glycosome. Proc. Natl. Acad. Sci U. S. A *102*, 11645-11650.

Wilkinson,S.R., Taylor,M.C., Horn,D., Kelly,J.M., and Cheeseman,I. (2008). A mechanism for cross-resistance to nifurtimox and benznidazole in trypanosomes. Proc. Natl. Acad. Sci. U. S. A *105*, 5022-5027.

Wilkinson,S.R., Taylor,M.C., Touitha,S., Mauricio,I.L., Meyer,D.J., and Kelly,J.M. (2002b). TcGPXII, a glutathione-dependent *Trypanosoma cruzi* peroxidase with substrate specificity restricted to fatty acid and phospholipid hydroperoxides, is localized to the endoplasmic reticulum. Biochem. J. *364*, 787-794.

Wilkinson,S.R., Temperton,N.J., Mondragon,A., and Kelly,J.M. (2000b). Distinct mitochondrial and cytosolic enzymes mediate trypanothione-dependent peroxide metabolism in *Trypanosoma cruzi*. J. Biol. Chem. *275*, 8220-8225.

Wilkinson,S.R., Horn,D., Prathalingam,S.R., and Kelly,J.M. (2003). RNA Interference Identifies Two Hydroperoxide Metabolizing Enzymes That Are Essential to the Bloodstream Form of the African Trypanosome. J. Biol. Chem. *278*, 31640-31646.

Wilkinson,S.R., Obado,S.O., Mauricio,I.L., and Kelly,J.M. (2002c). Trypanosoma cruzi expresses a plant-like ascorbate-dependent hemoperoxidase localized to the

endoplasmic reticulum. Proceedings of the National Academy of Sciences *99*, 13453-13458.

Wishart,D.S. and Sykes,B.D. (1994). The ^{13}C chemical-shift index: a simple method for the identification of protein secondary structure using ^{13}C chemical-shift data. Journal of Biomolecular NMR *4*, 171-180.

Wishart,D.S., Sykes,B.D., and Richards,F.M. (1991). Relationship between nuclear magnetic resonance chemical shift and protein secondary structure. Journal of Molecular Biology *222*, 311-333.

Wood,Z.A., Schroder,E., Robin,H.J., and Poole,L.B. (2003). Structure, mechanism and regulation of peroxiredoxins. Trends Biochem. Sci. *28*, 32-40.

Wuthrich, K. NMR studies of structure and function of biological macromolecules (Nobel Lecture). 2002.

Ye,Y. and Godzik,A. (2005). Multiple flexible structure alignment using partial order graphs. Bioinformatics *21*, 2362-2369.

Zhang,Y., Bond,C.S., Bailey,S., Cunningham,M.L., Fairlamb,A.H., and Hunter,W.N. (1996). The crystal structure of trypanothione reductase from the human pathogen *Trypanosoma cruzi* at 2.3 A resolution. Protein Sci *5*, 52-61.

APPENDIX

Table i: NMR resonance list for oxidised TcGPXI

Residue		N	HN	CA	CB	CO	HA1	HA2	NE1	HE1
14	**K**	118.294	8.904	56	34.7	175	4.53			
15	**S**	113.297	7.287	57.3	67.4	173.4	4.72			
16	**I**	113.064	9.22	64.4	38.9	174.4	3.98			
17	**Y**	115.294	7.339	59.5	36.6	175.2	4.45			
18	**E**	116.977	6.897	56.6	29.8	173.8	4.12			
19	**F**	117.788	7.324	58.3	41.7	172.7	4.66			
20	**Q**	120.461	8.917	55.7	30.9	174.7	4.87			
21	**V**	117.534	9.097	59.4	36.8	173.3	4.97			
22	**N**	119.838	8.698	52	40	174.9	5.42			
23	**A**	124.419	9.134	51.8	20.6	176.8	4.76			
24	**A**	121.021	7.949	55.9	18.8	177.2	3.63			
25	**D**	114.68	7.885	53.8	39.8	176.1	4.44			
26	**G**	108.535	8.483	45.3	-	172.8	3.66	4.296		
27	**K**	121.636	7.857	54.4	31.8	173	4.66			
28	**P**	-	-	64.3	32.5	175.6	4.84			
29	**Y**	126.497	9.379	57.1	41.5	174.6	4.66			
30	**D**	127.314	8.302	52.7	39.1	175.3	4.78			
31	**L**	130.09	8.421	58.1	43.7	176.8	4.33			
32	**S**	110.802	8.33	61.8	62.9	174.7	3.86			
33	**Q**	122.072	7.141	57.7	29.1	174.9	4.18			
34	**H**	118.82	7.742	56	29.4	172.8	4.72			
35	**K**	119.932	7.019	58.8	32.4	176.3	3.87			
36	**G**	113.127	7.946	44.9	-	172.4	3.54*			
37	**H**	119.133	8.292	53.1	31.3	170.1	5.18			
38	**P**	-	-	62.7	33.6	173.6	4.99			
39	**L**	113.753	8.386	55.6	45.5	173.4	4.95			
40	**L**	121.451	7.656	53.6	48.5	174	5.53			
41	**I**	122.179	8.897	61.5	40.6	175.3	4.97			
42	**Y**	126.72	9.162	57.7	43.5	174.3	5.44			
43	**N**	116.472	8.003	52.1	39	173.9	5.91			
44	**V**	115.892	8.539	58.4	35	173.7	4.08			
45	**A**	124.209	7.363	52.4	20.3	176.3	4.62			
46	**S**	115.306	8.268	59.4	63.7	-	4.39			
47	**R**	121.76	8.075	56	32.1	174.8	4.6			
48	**C**	-	-	56.4	42	173.7	4.62			
49	**G**	109.639	8.405	45.8	-	172.6	3.95*			
50	**Y**	120.128	8.086	58.4	39.3	174.6	4.58			
51	**T**	115.46	7.819	61.6	70	173	4.24			
52	**K**	122.787	7.919	56.9	33.1	176.1	4.26			
53	**G**	110.604	8.635	46.3	-	174.3	3.76	3.99		
54	**G**	110.767	8.1	46.7	-	173	3.12	3.70		
55	**Y**	119.06	7.632	61.4	42.8	176.9	3.93			
56	**E**	122.01	8.151	60.2	29.4	177.9	3.98			
57	**T**	120.002	8.173	66.7	68.8	174	3.66			
58	**A**	122.892	7.986	55.9	19.3	178.2	3.82			
59	**T**	112.625	8.134	67.7	69.3	175.5	3.98			
60	**T**	120.03	8.508	67.3	68.6	176	3.98			
61	**L**	120.67	8.296	58.6	42.8	177.1	3.92			

Table i: continued

Residue		N	HN	CA	CB	CO	HA1	HA2	NE1	HE1
62	Y	118.329	8.7	62.6	39	175.4	4.29			
63	N	114.339	8.418	56.6	38.8	176.6	4.16			
64	K	119.123	8.017	59.8	34.3	177.8	3.83			
66	K	123.834	8.574	59.3	30.1	179	4.88			
67	G	110.469	8.897	46.3	-	174.2	3.85	4.03		
68	Q	117.648	7.907	55.6	30.7	174.8	4.64			
69	G	107.382	7.859	45.1	-	171.7	3.45	3.60		
70	F	118.774	6.81	57.4	41.9	173.2	4.98			
71	T	118.86	6.232	60.9	70.6	169.2	4.45			
72	V	121.862	6.826	59.9	35.3	172.1	4.66			
73	L	120.203	8.167	52.9	45.5	174.8	4.87			
74	A	122.185	7.904	49.7	22.7	173.9	5.35			
75	F	122.783	9.434	55.1	39.9	172.4	4.6			
76	P	-	-	62	32.3	174.3	4.58			
77	C	119.213	8.442	58.1	29.9	174.1	4.63			
78	N	-	-	53.1	39.6	175.9	5.19			
79	Q	116.784	7.696	59	29.9	173.8	3.86			
80	F	118.396	8.147	56.5	38	174.5	4.74			
81	A	124.451	7.669	53.4	19.7	177.5	4.37			
82	G	108.308	8.558	45.5	-	173.1	4.01	4.02		
83	Q	119.316	8.139	55.9	30	174.9	4.41			
84	E	123.275	8.42	54.7	29.9	173.7	4.55			
85	P	-	-	63.9	32.3	176.6	4.39			
86	G	109.314	8.545	45.7	-	173.6	3.94	4.02		
87	T	112.813	7.852	62	70.4	173.5	4.35			
88	A	125.724	8.315	52.9	19.3	176.7	4.32			
89	L	120.516	8.062	55.7	42.6	176.5	4.28			
90	E	120.939	8.247	57	30.4	175.5	4.26			
91	V	120.771	8.004	62.7	32.8	175.3	4.06			
92	K	124.161	8.227	56.8	33.1	175.6	4.27			
93	E	120.702	8.24	57.1	30.4	175.2	4.2			
94	F	120.052	8.026	57.7	39.8	174.4	4.6			
95	A	124.535	8.152	52.6	19.4	176.2	4.31			
96	C	118.91	8.295	56.3	43.3	170.1	4.7			
97	T	-	-	62.4	70.2	173.3	4.4			
98	R	123.118	8.181	56.3	31.1	174.8	4.33			
99	F	123.603	8.307	59.1	39.8	173.2	4.22			
100	K	126.631	7.524	55.2	34.4	172.9	4.2			
101	A	124.323	7.768	51.4	21.3	175.2	4.07			
102	D	117.004	8.458	55.1	42	174.1	4.49			
103	F	115.653	5.823	52.5	36.7	171.6	4.48			
104	P	-	-	65	33.2	173.8	4.1			
105	I	121.314	7.506	59.5	39.2	175.4	4.62			
106	M	125.613	8.385	53	35.7	174.3	4.62			
107	A	119.059	7.551	52.8	19	175.8	4.1			
108	K	119.452	7.945	56.9	32.8	175.8	4.53			
109	I	118.931	8.405	60	44	175	4.81			
110	D	119.796	8.555	54.4	41.1	174.7	4.85			
111	V	112.032	7.919	60.7	33.1	173	4.13			

Table i: continued

Residue		N	HN	CA	CB	CO	HA1	HA2	NE1	HE1
112	**N**	117.852	8.1	51.3	43.4	175.3	4.82			
113	**G**	-	-	-	-	-	-			
114	**S**	-	-	61.6	63.5	174.5	4.1			
115	**K**	120.293	8.572	54.5	31.5	174.7	4.54			
116	**A**	122.526	7.127	52	18.8	176.3	3.51			
117	**H**	124.463	8.169	57	33	174.9	4.44			
118	**P**	-	-	65.7	32.6	177.2	4.09			
119	**L**	124.24	11.465	58.5	42.3	178.6	2.7			
120	**Y**	117.859	7.41	64	38.9	177.1	3.38			
121	**E**	118.005	7.623	59.3	29.1	178.1	4.25			
122	**F**	119.222	7.47	60	39.6	177	4.4			
123	**M**	122.023	7.803	61.7	31.4	175.6	2.78			
124	**K**	115.148	8.345	59.2	33.2	175.7	3.67			
125	**A**	113.864	7.015	53.4	18.8	178.4	3.88			
126	**T**	114.402	7.253	66.7	68.5	173.1	3.73			
127	**I**	118.115	7.497	56.7	37.4	171.3	4.5			
128	**P**	-	-	63.2	32.6	175.2	4.5			
129	**G**	105.966	8.014	44.2	-	172.8	3.85	4.01		
130	**L**	121.226	8.131	56.6	42.3	177.6	3.95			
131	**F**	117.998	8.776	58.1	37.4	175.3	4.49			
132	**G**	106.751	8.226	46	-	173.7	3.60	4.21		
133	**T**	114.558	7.393	61.8	70	174.1	4.35			
134	**K**	-	-	59.6	33.7	176.8	3.98			
135	**A**	120.279	7.824	53.7	19.3	176.9	4.02			
136	**I**	122.412	8.548	59.9	36.6	176.9	-			
137	**K**	127.141	9.16	56.8	33.4	173.8	3.97			
138	**W**	112.468	6.163	55.4	34.1	176.4	4.55		130.57	10.26
139	**N**	120.899	8.68	54.4	39.8	177.7	4.14			
140	**F**	114.257	8.574	62.2	35.8	175.4	4.2			
141	**T**	116.523	7.204	65.7	69.6	172	4.37			
142	**S**	119.145	8.535	56.7	66.2	172.3	5.53			
143	**F**	115.492	8.473	56.3	44	174.1	5.37			
144	**L**	124.725	8.653	54	44.9	172.5	5.05			
145	**I**	130.29	9.603	57.9	37.6	174.4	4.57			
146	**D**	123.954	8.491	53.1	41.7	177.6	4.6			
147	**R**	113.845	8.572	58.8	30	174.7	3.75			
148	**H**	118.736	8.539	55.5	30.6	175.1	4.82			
149	**G**	107.003	8.234	47.2	-	172.2	3.42	4.34		
150	**V**	126.014	8.851	60.6	33.3	173.5	4.53			
151	**P**	-	-	62.7	31.5	174.8	4.94			
152	**V**	115.257	9.331	61.6	34.5	174.9	4.53			
153	**E**	117.626	7.4	55.3	34	171.9	4.54			
154	**R**	126.558	8.841	54.2	34	173.5	4.7			
155	**F**	126.662	9.583	57.2	40.1	174.7	4.85			
156	**S**	118.597	8.8	57.1	63.2	173.2	4.75			
157	**P**	-	-	64.7	32	175.2	4.58			
158	**G**	106.482	8.004	44.8	-	173.2	3.58	4.03		
159	**A**	124.182	7.294	53	19.1	176.1	4.34			
160	**S**	120.128	9.08	56.6	66.6	173.7	4.53			

Table i: continued

Residue		N	HN	CA	CB	CO	HA1	HA2	NE1	HE1
161	**V**	119.789	8.518	67.3	32.4	176.4	3.29			
162	**E**	117.697	8.442	60.5	29.4	178	4.13			
163	**D**	120.573	8.029	57.6	40.7	177.8	4.23			
164	**I**	119.167	7.463	64.7	37.6	178.5	3.33			
165	**E**	121.365	9.105	60.9	30.3	176.4	3.53			
166	**K**	115.643	7.568	59.7	33	177.2	3.98			
167	**K**	113.015	7.06	55.3	33.9	176.2	4.44			
168	**L**	120.778	8.671	57.4	44.1	176.5	4.08			
169	**L**	118.696	8.539	60.4	38.7	176.9	3.92			
170	**P**	-	-	65.7	31.8	177.6	4.36			
171	**L**	114.554	7.325	55.9	42.5	177.9	4.34			
172	**L**	119.103	7.676	56	42.8	177	4.28			
173	**G**	106.547	7.929	45.7	-	173.6	4.03*			
174	**G**	108.408	8.111	45.3	-	172.7	3.98	3.975		
175	**A**	123.386	8.16	52.5	19.8	176.5	4.35			
176	**R**	121.097	8.328	56.2	30.9	174.4	4.38			
177	**I**	126.502	7.716	63	39.9	180.1	4.09			

(* denotes degenerate glycine HA values)

Table ii: NMR resonance list for reduced TcGPXI

Residue		N	HN	CO
14	K	118.366	8.891	175.01
15	S	113.25	7.335	173.393
16	I	112.377	9.19	174.45
17	Y	115.232	7.318	175.242
18	E	116.584	6.906	173.965
19	F	117.658	7.33	172.768
20	Q	120.592	8.907	-
21	V	117.164	9.024	173.996
22	N	119.792	8.669	174.425
23	A	123.101	9.029	177.023
24	A	120.968	7.971	176.496
25	D	115.825	7.938	176.042
26	G	108.716	8.526	173.016
27	K	121.867	7.903	-
28	P	-	-	175.428
29	Y	126.526	9.256	174.351
30	D	127.251	8.27	175.361
31	L	129.806	8.535	176.922
32	S	111.241	8.377	173.54
33	Q	121.805	7.14	174.955
34	H	118.662	7.775	172.749
35	K	119.869	7.031	176.382
36	G	112.989	7.9	172.478
37	H	119.511	8.349	-
38	P	-	-	173.683
39	L	113.929	8.418	-
40	L	-	-	-
41	I			175.482
42	Y	126.546	9.023	-
43	N	-	-	-
44	V	-	-	-
45	A	-	-	-
46	S	-	-	-
47	R	-	-	-
48	C	-	-	-
49	G	-	-	-
50	Y	-	-	-
51	T	-	-	-
52	K	-	-	-
53	G	-	-	-
54	G	-	-	-
55	Y	-	-	-
56	E	-	-	-
57	T	-	-	-
58	A	-	-	-
59	T	-	-	-
60	T	-	-	176.161
61	L	120.456	8.267	-

Table ii: continued

Residue		N	HN	CO
62	Y	-	-	-
63	N	-	-	176.966
64	K	118.941	7.998	177.888
65	Y	112.006	7.548	176.853
66	K	124.19	8.82	-
67	G	-	-	174.24
68	Q	117.716	7.943	174.8
69	G	107.21	7.887	171.745
70	F	119.057	6.899	173.234
71	T	119.054	6.299	169.403
72	V	122.589	6.504	172.304
73	L	120.233	8.195	174.617
74	A	122.492	7.891	173.828
75	F	122.987	9.434	-
76	P	-	-	-
77	C	-	-	-
78	N	-	-	-
79	Q	-	-	-
80	F	-	-	-
81	A	-	-	-
82	G	-	-	-
83	Q	-	-	-
84	E	-	-	-
85	P	-	-	-
86	G	-	-	-
87	T	-	-	-
88	A	-	-	-
89	L	-	-	-
90	E	-	-	-
91	V	-	-	-
92	K	-	-	-
93	E	-	-	-
94	F	-	-	-
95	A	-	-	-
96	C	-	-	-
97	T	-	-	-
98	R	-	-	-
99	F	-	-	-
100	K	-	-	-
101	A	-	-	-
102	D	-	-	174.535
103	F	116.521	5.932	-
104	P	-	-	-
105	I	-	-	175.454
106	M	126.819	8.438	-
107	A	-	-	175.839
108	K	119.098	7.902	176.571
109	I	118.57	8.413	174.837

Table ii: continued

Residue		N	HN	CO
110	**D**	119.838	8.562	174.446
111	**V**	112.185	7.96	-
112	**N**	-	-	-
113	**G**	-	-	-
114	**S**	-	-	174.48
115	**K**	120.325	8.594	174.734
116	**A**	122.469	7.131	176.359
117	**H**	124.401	8.194	-
118	**P**	-	-	177.266
119	**L**	123.86	11.395	178.719
120	**Y**	117.737	7.459	177.191
121	**E**	118.078	7.621	178.071
122	**F**	119.133	7.475	176.963
123	**M**	121.799	7.824	175.835
124	**K**	115.317	8.371	175.684
125	**A**	114.121	7.066	178.453
126	**T**	114.167	7.27	173.215
127	**I**	118.144	7.503	-
128	**P**	-	-	-
129	**G**	-	-	-
130	**L**	-	-	-
131	**F**	-	-	175.214
132	**G**	107.033	8.263	-
133	**T**	-	-	-
134	**K**	-	-	176.935
135	**A**	120.284	7.898	-
136	**I**	-	-	-
137	**K**	-	-	-
138	**W**	-	-	-
139	**N**	-	-	-
140	**F**	-	-	-
141	**T**	-	-	171.895
142	**S**	119.215	8.549	-
143	**F**	115.881	8.564	174.02
144	**L**	124.773	8.784	172.637
145	**I**	129.991	9.647	174.356
146	**D**	124.103	8.54	177.593
147	**R**	113.843	8.532	174.562
148	**H**	118.904	8.539	175.176
149	**G**	107.164	8.268	172.251
150	**V**	125.571	8.848	-
151	**P**	-	-	174.648
152	**V**	115.575	9.302	174.988
153	**E**	117.7	7.459	172.107
154	**R**	126.259	8.764	173.177
155	**F**	126.935	9.601	174.594
156	**S**	118.625	8.778	-
157	**P**	-	-	-

163	**D**	120.302	7.954	-
164	**I**	-	-	178.559
165	**E**	121.356	9.2	176.494
166	**K**	115.206	7.496	177.194
167	**K**	113.02	7.09	176.184
168	**L**	120.923	8.574	176.599
169	**L**	118.672	8.543	-
170	**P**	-	-	177.662
171	**L**	114.826	7.38	177.985
172	**L**	118.959	7.71	177.018
173	**G**	106.487	7.95	173.647
174	**G**	108.412	8.11	172.734
175	**A**	123.317	8.156	176.486
176	**R**	120.982	8.318	174.404
177	**I**	126.398	7.706	-